1/2 ベージ Searching PAJ

¥

PATENT ABSTRACTS OF JAPAN

10-116094 (11)Publication number:

(43)Date of publication of application: 06.05.1998

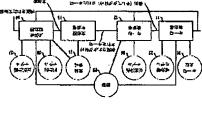
(71)Applicant: LUCENT TECHNOL INC JUANG BIING HWANG (72)Inventor: UU CHOU G10L 3/00 G10L 3/00 (21)Application number: 09-265959 30.09.1997 (22)Date of filing: (51)Int.Cl

KAWAHARA TATSUYA CHIN-FUI LEE

Priority country : US US Priority date: 01.10.1996 20.12.1996 Priority number: 96 724413 96 771732 (30)Priority

(54) METHOD AND DEVICE FOR VOICE RECOGNITION

SOLUTION: A key phrase detector 11 detects plural key phrases based on the set of the phrase part grammar of phrases, the scales are compared with threshold values phrase verifier 12 assigns reliability scales to these key the intrinsic condition of the conversation. Then, a key hypothesis generator 13 couples the candidates based conversation based on a fixed and formalized grammar. information 25 and a sentence hypothesis is obtained. freedom while accepting freer conversation than the on a prescribed (task intrinsic, for example) meaning conversation voice recognition with a high degree of PROBLEM TO BE SOLVED: To efficiently conduct to verify the key phrases and a set of verified key phrase candidates is obtained. Then, a sentence



LEGAL STATUS

the conversation in a large scale task (i.e., against 'subtasks').

grammar is used in accordance with the specific state of

generated and a final recognition result is obtained. In

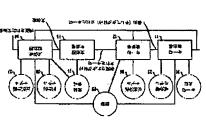
sentence hypothesis, verified sentence hypothesis is the conversation base system, individual phrase part

Lastly, a sentence hypothesis verifier 14 verifies the

05.04.1999 [Date of request for examination]

[Date of sending the examiner's decision of

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]



3 榖 (12) 公開特許公 (19)日本国特許庁 (JP)

特開平10-116094

(11)特許出顧公開番号

(43) 公開日 平成10年(1998) 5月6日

561G 531C 3/00 G10L 数列配中 561 5 3 1 3/00

> G10L (51) Int.Cl.

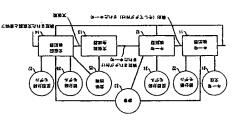
審査闘求 末髄求 韻求項の数32 OL (全 12 頁)

(21) 出版器号	特閣平9-265959	(71) 出國人 596077259	586077259
			ルーセント テクノロジーズ インコーボ
(22) 出版日	平成9年(1997)9月30日		レイデッド
			Lucent Technologies
(31)優先権主強番号	(31)優先権主強番号 08/724413		Inc.
(32)優先日	1996年10月1日		アメリカ合衆国 07974 ニュージャージ
(33)優先格主張国	米取 (NS)		一、マレーヒル、マウンデン アベニュー
(31) 優先権主張番号	08/771732		000 – 200
(32)優先日	1996年12月20日	(72) 発明者	(72)発明者 ウー チョウ
(33)優先指主頭国	米国 (ns)		アメリカ台衆国、07922 ニュージャージ
			ー、パークレー ハイツ、グリーンブライ
			アー ドライブ 22
		(74)代理人	(74)代理人 弁理士 三段 弘文
			最終頁に被《

音声認識方法および音声認識装置 (54) [発明の名称]

よりも多くの発話が受容される、効率および自由度の高 【課題】 固定した形式的な文法に基づいて受容される い会話音声認識を実現する。

る。次に、キー句後証器12で、これらのキー句に信頼 句候補の集合を得る。次に、文仮脱生成器13で、検証 済みキー句候補を、所定の(例えばタスク固有の)意味 情報25に基づいて結合して文仮説を得る。最後に、文 仮散検証器14で、これらの文仮説を検証して、検証済 み文仮説を生成し、最終認識結果を得る。全話ベースの システムでは、大規模なタスク内の会話の特定の状態に 応じて(すなわち「サブタスク」に対して)、個別の句 【解決手段】 キー句検出器11で、会話の状態に固有 性尺度を割り当て、その信頼性尺度をしきい値と比較す ることによってそれらのキー句を検証し、検証済みキー の句部分文法の集合に基づいて複数のキー句を検出す 部分文法を使用することが可能である。



<u>|</u>

http://www19.ipdl.ncipi.go.jp/PA1/result/detail/main/wAAArbaiNwDA410116094P... 2007/02/27

特開平10-116094 3

複数の単語からなる音声発話の音声認識 を実行する音声認識方法において、

句部分文法に基づいてキー句検出を奥行して、認識され た単語からなる複数の検出済みキー句を生成する検出ス 前記検出済みキー句に信頼性尺度を割り当て、駭信頼性 尺度をしきい値と比較することにより、前配検出済みキ 一句の検証を実行して、検証済みキー句候補の集合を生 成するキー句検証ステップと 前記検証済みキー句候補を結合し、所定の意味情報に基 ろいて文仮説を生成するステップと、

帕記文仮説の検証を実行して、少なくとも1つの検証済 み文仮説を生成する文仮説検証ステップとからなること を特徴とする音声認識方法。

【榊求項2】 前配句部分文法は、会話状態に基づく句 部分文法の集合から選択されることを特徴とする請求項

の方法。

いえを用いたトレーニングプロセスに基づいて導出され

20

(請求項4) 前記文仮説の生成は、前記信頼性尺度に たものであることを特徴とする請求項1の方法。 も基づくことを特徴とする請収項1の方法。 【精水項5】 前記検出済みキー句は、意味タグでラベ 【酵水項6】 前記文仮説の生成は、前記意味タグにも ルされることを特徴とする開東項1の方法。

基づくことを特徴とする請求項5の方法。

【請求項7】 前配文仮説の生成は、前記信頼性尺度に も基づくことを特徴とする請求項6の方法。

最も確からしい文仮説を判定するステップからなること 「請求項8」 前記文仮説の生成は、前記信頼性尺度、 前記意味タグ、および前記所定の意味情報に基づいて、 を特徴とする開水項7の方法。

30

に基づいて実行され、前記検出済みキー句は部分語の列 【請求項9】 前記後出ステップは複数の部分語モデル からなることを特徴とする鯖末頃1の方法。

【精氷項10】 前配部分話モデルは隠れマルコフモデ ルからなることを特徴とする請求項9の方法。

デルの集合と、対応する反部分語モデルの集合に基ろい 【精水項11】 前記キー句検証ステップは、部分語モ て実行されることを特徴とする開東項9の方法。

40

【韓末項12】 前配部分語モデルおよび前配反部分語 モデルは隠れマルコフモデルからなることを特徴とする 構束項11の方法。

はして音響的検証を実行するステップからなることを特 【精求項13】 前記文仮説検証ステップは、文仮説に 徴とする請求項1の方法。

[請求項14] 前記文仮設後証ステップは、文仮説に **対して意味的検証を実行するステップからなることを特**

【請求項15】 前記文仮説検証ステップは、最も確か らしい 1 つの文仮説を違択するステップを含むことを特 徴とする請求項1の方法。

レームを生成するステップをさらに有することを特徴と 【精水項16】 前記検証済み文仮説に基ろいて意味フ する請求項1の方法。

【請求項17】 複数の単語からなる音声発話の音声器 歳を実行する音声路職装置において、

た単語からなる複数の検出済みキー句を生成するキー句 句部分文法に基づいてキー句検出を実行して、認識され 極田器と、

0/

前記検出済みキー句に信頼性尺度を割り当て、該信頼性 尺度をしきい値と比較することにより、前記検出済みキ 一句の検証を実行して、検証済みキー句候補の集合を生

成するキー句検証器と、

前配検証済みキー旬候補を結合し、所定の意味情報に基 づいて文仮説を生成する文仮説生成器と、 前記文仮説の検証を実行して、少なくとも1つの検証券 み文仮説を生成する文仮説検証器とからなることを特徴 とする音声認識装置。

【讃求項18】 前配句部分文法は、会話状態に基づく 句部分文法の集合から選択されることを特徴とする請求

ーパスを用いたトレーニングプロセスに基ろいて導出さ 【鯖水頂19】 前記句部分文法は、音声サンプルのコ れたものであることを特徴とする請求項17の装置。

【請求項20】 前配文仮說生成器は、前記值類性尺度 にも基づいて前記文仮説を生成することを特徴とする詩 末項17の装置。

【精水項22】 前記文仮説生成器は、前記意味タグに [請求項21] 前記検出済みキー句は、意味タグでラ も基づいて前記文仮説を生成することを特徴とする請求 ベルされることを特徴とする間求項17の装置。 頃21の装置。

にも基づいて前記文仮説を生成することを特徴とする詩 【請求項23】 前記文仮説生成器は、前記情頼性尺度 求項22の装置。

て、最も確からしい文仮説を判定することを特徴とする 度、前記意味タグ、および前記所定の意味情報に基づい 【請求項24】 前記文仮說生成器は、前記信頼性尺 請求項23の装置。

ルに基づいて動作し、前記後出済みキー句は部分語の列 【請求填25】 前記キー句検出器は複数の部分脳モデ からなることを特徴とする請求項17の装置。 【請求項26】 前配部分語モデルは隠れマルコフモデ

集合と、対応する反部分語モデルの集合に基づいて動作 【請求項27】 前記キー句検証器は、部分語モデルの ルからなることを特徴とする請求項25の装置。 することを特徴とする請求項25の装置。 【糖水項28】 前記部分語モデルおよび前記反部分語

20

散とする請求項1の方法。

-2-

モデルは隠れマルコフモデルからなることを特徴とする 請求項27の装置 [精水項29] 前記文仮散検証器は、文仮説に対して 音響的検証を実行することを特徴とする請求項17の装 [請求項30] 前記文仮設検証器は、文仮説に対して 意味的検証を実行することを特徴とする請求項17の装

1 つの文仮説を遊択することを特徴とする請求項 1 7 の [請求項31] 前記文仮散後証器は、最も確からしい

[請求項32] 前記検証済み文仮脱に基ろいて意味フ レームを生成する意味フレーム生成器をさらに有するこ とを特徴とする請求項17の装置。

[発明の詳細な説明] [0001]

[発明の属する技術分野] 本発明は、音声閣職の分野に 関し、特に、会話音声を理解する方法に関する。

[0002]

【従来の技術】過去数年間、会話音声の認識および理解

のためのシステムが開発され、いくつもの「実世界」の 応用で評価されている。いくつかのアプローチが用いら 一ションに限定されるが、ユーザの発話を受け取る(そ る。このようなシステムでは、認識器は、音声入力全体 れている。第1のアプローチは決定性有限状態文法(F SG)である。これは、簡単なクスクあるいはアプリケ してそれにより認識し最終的には理解する)ものであ を、可能な(すなわち、固定した文法により許容され ろ) 単語列のいずれかに合うものを見つける (彼号す ろ)ことを試みる。

語、ロごもり、反復および予想外の投現などがある。日 タスクの場合でさえ、自然なユーザ発話の20%以上が のような複雑な問合せに対処する固定したタスクベース の文法を書いてから十分に閲覧することは、注外な畳の 固定した文法を使用することは、ほとんど文法的に正し の典型的な「実世界」の環境では、さまざまな発話に遺 過し、その多くはこのようなタスクベースの文法によっ て十分に対応することができない。このような文法的に 文法外となる可能性が高いことが分かっている。このよ うな条件下で、これらの文法を用いるシステムの性能は 低い。そしてこの低い性能は、試行期間中に文法を念入 りに腐敗したにもかかわらず起こる。タスクがさらに複 雑な問合せに関わる場合には状況はさらに悪くなる。こ [発明が解決しようとする課題] 実際には、このような い文 (文法内の文) がシステムに与えられる場合には有 効である。しかし、多数のユーザに適用されるほとんど 付あるいは時刻の音声認識のような明らかに単純なサブ 正しくない文(文法外の文)には、例えば、無関係単

40

特開平10-116094

ල

する)「フィルタ」モデルの使用は、固定文法のほとん 号の枠組みから生じている。 (発話の文法外部分に適合 が、固定文法に固有の基本的な問題点を解決していない [0004]上記の問題点は、固定した文レベルの文法 しなければならないという一様な要求条件を適用する復 を仮定して、その文法が入力全体に適合(マッチング) ど従う音声サンプルには限定された成功を収めている ために、多くの一般的な発話が認識されないままであ

つのアプローチは、統計首語モデルの使用に関するもの である。このようなモデルは、固定した所定の文法に基 (トレーニング) の結果として統計的に生成された文法 となっているATIS(航空旅行情報システム(Air Tra モデルのアプローチを用いた会話音声処理に対する包括 的プロジェクトである。(例えば、D. A. Dahl, "Expan **ータが収集され、文法外の発話を認確する能力に関して** に基づく。倒えば、米国政府(ARPA)がスポンサー 【0005】会話音声の認識および理解に対するもう1 ding the Scope of the ATIS Task: The ATIS-3 Corpu vel information System)) プロジェクトは、統計書語 s". Proc. ARPA Human Language Technology Workshop. pp. 43-48、1994、参照。)この場合、かなりの量のデ ろくのではなく、大量のサンプルデータを用いた学習 も、統計宮脇モデルの使用は比較的成功を収めた。

20

にもかかわらず、実際には上記のような決定性有限状態 ションでは、データ収集作業自体が大量の (人間の) 時 聞および労力を必要とするため、統計モデルをトレーニ ングするためにこのような大量のデータを提供すること ンでは、与えられたタスクに対して単純な2連語(bigra n) 含語モデルを構築するのに十分な量のデータを収集す るのでさえ実際的ではない。 (さらに、注意すべき点で あるが、ATISシステムの場合、実行されたシナリオ およびデータ収集はやや人工的であり、従って、実世界 の自然発話に固有の問題点を必ずしも反映していない可 能性がある。)これらの理由から、「実世界」のアプリ ケーションに配備され試験されているほとんどの会話シ ステムは、文法外発話を処理する能力が制限されている は実際的でないことが多い。ほとんどのアプリケーショ 【0006】しかし、一般的な「実世界」のアプリケー

30

Rose, "Keyword Detection in Conversational Speech U tterances Using Hidden Markov Model Based Continuo ge, 9(9):309-333, 1995、および、H. Tsuboi and Y. I [0007] 会話音声認識問題のために考えられている us Speech Recognition", Computer Speech and Langua akebayashi, "A Real-time Task-oriented Speech Unde rstanding System Using Keyword-spotting", Proc. 1E グ」方式に関するものである。これは、例えば、R. C. もう1つのクラスのアプローチは「単語スポッティン 文法を使用している。 30

(人間の) 時間および労力を必要とすることが明らかに

€

特開平10-116094

EE-IGASSP, volume 1, pp. 197-200, 1992、に記載されている。これちのアプローチは、入力発脂の非キーワーでいる。 まま物のモデルに使用する方法に依存して2つのカテード・ペッキュー

2, pp. 459-462, 1993、および、M. Weintraub, "Keywo 固有の問題点を解決していない。 さらに、大語彙自然音 きる単語スポッティング方式は、大語彙を認識する能力 に基づくものである。この方式の例としては、J. R. Ro licek et al., "Phonetic Training and Language Wod eling for Word Spotting", Proc. IEEE-ICASSP, volume Speech-RecognitionSystem", Proc. IEEE-ICASSP, volu み、キーワードモデルとともに、多くの非キーワード語 彙単語モデルをシステムで利用可能とする。しかし、こ の技術でも、自然音声でしばしば見られるロごもりや自 己反復のような非適格な現象を十分にモデル化すること ができない。すなわち、すべての一様復号アプローチに 【0008】その第1のカテゴリーに分類することがで rd-Spotting Using SRI's DECIPHER Large-Vocabulary pp.463-466, 1993, に記載されているものがあ る。この技術は、できるだけ多くの語彙知識を組み込 **声認識技術は、タスク領域が限定される場合には特に、** 性能および効率性に問題がある。

peech Using Hidden Markov Models", IEEE Trans. Aco モデルの並列ネットワークも、非キーワードに合うもの を見つける性能が十分ではなく、そのため、キーワード 「認識」) が起こる。さらに、このカテゴリーに属する 部分語に基づく音声影識アプローチの利点の多くを 徴牲にしている。これらの理由から、このカテゴリーの 単語スポッティング方式で応用が成功しているのは、例 えば数字音声認識のタスクのような非常に小さい語彙を 【0009】単語スポッティング方式の第2のカテゴリ **ーは、入力発話の非キーワード部分をモデル化するため** クまたは単純ガーペジモデルとともに単純 (すなわち限 定語彙)単語スポッティングを使用するものである。こ のような方式の例としては、J. G. Wilpon et al.. "Au tomatic Recognition of Keywords in Unconstrained S 990、および、R. C. Rose and D. B. Paul, "A Hidden Markov ModelBased Keyword Recognition System", Pr るものがある。残念ながら、ガーベジモデルも、部分語 モデルが発話の無関係(すなわち非キーワード)部分と 誤って合わせられることも多い。この結果、多くの「フ **芫育のほとんどのシステムは、語彙に依存してキーワー** ドモデルおよびガーベジモデルを「糊整」し、それによ oc. IEEE-ICASSP, pp.129-132, 1990、に記載されてい に用いられる部分語(subword)モデルの並列ネットワー ust., Speech & Signal Process., 38(11):1870-1878, オールスアラーム」(すなわち、キーワードの嵌った 含むタスクのみである。

[0010]

[課題を解決するための手段] 本発明の発明者が認識し

分的あるいは全体的な理解が可能となる一方で、発話の は、本明細帯では、任意の単語列を意味し、そのような 単語列が文法的に正しい文構造を有するかどうかは問わ るアプローチに払づいて構築することができる。従来の 文法的制約を緩和し、かつ、認識されるキー句の集合に **怖に注目することによって、例えば固定した形式的な文** 句 (キーフレーズ) 」を含み、その踢簸により発話の部 ない。また、「キー句」という用語の使用は、本明細鸛 なわち、キーワードは単一の単語からなる「キー句」で 女の意味的に重要な部分を検出し無関係な部分を拒絶す 他の部分は実際にはタスクに関係がないので無視すべき では、1個以上の単語からなる列を含むものとする。す 「文」) はタスクに関するあるキーワードおよび「キー ある。)すなわち、自由度の高い音声理解システムは、 生に基づいて受容されるよりも多くの発話が受容され である。(注意すべき点であるが、「文」という用語 たところでは、ほとんどの会話音声発話(すなわち

【0011】そこで、本発明の実施例によれば、自由度の高い(すなわち、制約のない)音声の理解を実現するために使用可能なキー句の検出および検証の技術が実現さる。具体的には、単語列(すなわちズ)からなる音質を活し、多値パス1手続きが適用される。まず、例えば全話の状態に固有の句部分文法の集合に基づいて複数のキー句を検出(すなわち、路離)する。次に、これのキー句に横作尺度を割り当て、その循係住尺度とのキー句に横作尺度を割り当て、その循係住尺度といくその結果として、検証済みキー句候補の集合を得る。次に、検証済みキー句候補の集合を得る。次に、検証済みキー句候補の集合を得る。次に、検証済みキー句候補の集合を得る。次に、検証済みキー句候補を、所定の(例えばタスク監有の)意味情報に基づいて結合して文匠股を得る。文成記は、国々のキー句信頼性及に基づいて生成することも可能である。最後に、これらの文匠股を得距しまる。

20

[0012] さらに、会話ペースのシステムでは特に、大規模なタスク内の会話の特定の状態に応じて(すなわち、「サブタスク」に対して)、国別の句部分文法を使用することが可能である。例えば、会話ペースの目動車下が要えかかない。といいは、大・要はされて単が必要となる日時を決定する必要がある。この場合、予期される応答は、時間的な情報のみを与えるものであると限定することができる。自由度の高い会話マネージャと組み合わされることにより、本税同によるシステムは、文音声を少なくとも部分的に理解することができる。さらに、会話セッションが進むうちに、必要な明確化(機味さの除土)を実行することも可能である。

(0013)

【発明の実施の形態】 30 【はじめに】本発明の実施例によれば、会話音声の認識

および理解のためのシステムは、(例えば、非キーワード大語彙知識を用いることなく) 部分語ペースの音声路 職の一般的な枠組みで、無関係部分を関って「認識」せずに、発話の直要部分を認識することによって契奨される。(明分語ペースの音声認識は、当業者には周知であるが、音節、半音節あるいは音楽のような単語をグメントでなおよびでッチングを含む。次に、それらの単語セグメント(すなわち、部分語)の別に、疑察内の単語をランとして、単語に対応するモデルは、疑葉的に、辞書によって指定される、その単語を構成する部分に、辞書によって指定される、その単語を構成する部分に、許書によって指定される。との単語を構成する部分に、許書によって指定される。との単語を構成する部分による音声影響なおよび音音を発展のまる。)図1に、本発明の実施例による音声影響ないで音音を高います。

【0014】注意すべき点であるが、従来技術の最も直 に、その結果にどのくらいの信頼性をおくことができる かが分からないことである。この理由で、図1に示した 本発明の実施例によれば、認識した結果に対する仮脱検 定を実行し、それに信頼性尺度を割り当てる検証方法を 用いる。 (例えば、R. A. Sukkär et al., "A Vocabula ry Independent Discriminitively Trained Method for Recognition", Proc. EuroSpeech-95, pp. 1629-1632, 1 of Keyword Strings Using Word-Based Minimum Verifi P. pp.518—521, 1996、および、M. Rahím et al., "Dis criminitive Utterance Verification Using Ninimum S tring Verification Error (MSVE) Training". Proc. 1 EEE-ICASSP, 1996、参照。) このような発話検証法を図 1の実施例のシステムに統合することによって、キーワ 一ド(あるいは、この場合にはキー句)の検出の信頼性 を高くすることができる。すなわち、キーワードモデル への正しくないマッチングすなわち 「フォールスアラー 995, R. A. Sukkar et al., "Utterance Verification Rejection of Non-Kaywords in Subword-BasedSpeech cation Error (WB-MVE) Training", Proc. IEEE-1CASS 大な問題点のうちの1つは、従来の音声認識器は一般 ム」は大幅に減少する。

30

[0015]また、図1の実施例のシステムは、このような「フォールスアラーム」をさらに確少させる。システムは、このようなキーワード(あるいはキー句)マッテングおよび後証プロセスの単独の結果として「最終判定」をしたが、むしろ、後証したキーの一形あらいはキー句の紹み合わせ(すなわち、双)に基づいて意味解析を実行して大仮限を生成し、それを別の終証プロセスで、後証する、特に、この文仮解検証プロセスは、生発語内にあるいくつかの組分額からなる「部分入力」で実行さにあるいくつかの組分額からなる「部分入力」で実行さ

(100)6)既に指摘したように、図1の実施例のシステムは、検出単位として、キーワードのみを使用するのではなく、キーのを使用する。上記の単語スポッティング方式は一般に、局所的ノイスや気体な音によって容易

特開平10-116094

9

にトリガされる小さいテンプレートを使用する。より長い後出単位(すなわち、単なるキーワードの代わりにキー句)を使用することは、より特徴的な情報を含むことになり、その結果、疑難段略および検証段略の両方で、より安定な音響マッチングが得られるので、有効であ

。 【0017】具体的には、キー句は、1個以上のキーワードと、おそらくは、機能脂との列からなる。例え

直接対応するように定義される。(意味フレームは、当 される情報テンプレートからなる。) 従来のn 連語(n-g Language Models for Spontaneous Speech", Proc. 40 ams for Continuous Speech Recognition", Proc. 1EEE F. Bimbot, "Language Modelingby Variable Length Se 参照。)とは異なり、本実施例によって認識されるトッ 珠(セマンティック)フレームにおける意味スロットに ンに対して、会話によって部分的にあるいは完全に充填 3LP, pp. 831-834, 1994, E. P. Giachin, "Phrase Bigr quences: Theoretical Formulation and Evaluation of 際には、キー句は、例えば時刻および場所のような、意 業者に周知の用語であるが、与えられたアプリケーショ ram) 質酷モデルによって定義されるようなボトムアップ 句 (例えば、B. Suhm and A. Waibel, "Towards Better プダウンキー旬は、容易に意味表現へと直接にマッピン [0018]ここに記載する本発明の実施例によれば、 Multigrams", Proc. IEEE-ICASSP, pp.169-172, 1995, グされる。従って、これらのキー句の検出は、直接に、 -ICASSP, pp. 225-228, 1995、および、S. Deligne and 検出されたキー句には概念情報のタグが付けられる。

20

-句文法21)の集合を用いてキー句の集合を認識する られる。このタグは、文仮脱生成器13(後述)によっ は、辞書23および部分籍モデル22を使用する。これ 3および文仮脱検証器14を有する。特に、キー旬検出 器11は、会話状態に特有の句部分文法(すなわち、キ ための部分語ベースの音声総職器からなる。検出された キー句には、次に、意味(セマンティック)タグが付け てその後に実行される文レベルの解析で有用となる。キ 一句検出器11によって用いられる部分語モデル認識器 らは、例えば、当業者に周知の従来の最小分類開整(M CE(minimum classification error)) 基準に基づいて キー句検出器11、キー句検証器12、文仮脱生成器1 [0019] 具体的には、図1の実施例のシステムは、 トレーニングされたものである。これらのモデル自体 (HMM) からなることも可能である。 発話の確実な理解につながる。

[1020] 太に、検出されたキー句は、キー句検証器

-5-

30

特理平10-116094

込む必要がある。 VE(minimm verification error)) 基準を用いてトレ 2 は辞售23、部分階モデル22および反朗分簡モデル 2 4 を使用する。これらは、例えば、最小検証関整(M 上記のように、このプロセスは、これがなければ超こり 得る多くのフォールスアラームを除去する。奥施例のキ 部分語レベルの検証の担合せからなる。キー句検証器1 一句検証器12は、当業者に周知の「反部分語モデル」 を用いて、認識されたキー句の各部分語をテストする、 12によって検証され、債頼性尺度が割り当てられる。 ーニングされたものである。

【0021】図1の実施例の第3の構成要素は文仮脱生 成器 1 3 である。これは、例えばタスク固有の意味情報 説へと結合する。例えば、T. Kawahara et al., "Conce Speech Understanding", Proc. IEEE-ICASSP, pp.291-2 25を用いて、検証されたキー句候補を1つ以上の文仮 て、意味制約を満たす最適な仮説を探索することができ pt-Based Phrase Spotting Approach for Spontaneous 94, 1996、に記載されたようなスタック復号器を用い

【0022】最後に、文仮説検証器14によって、音響 (すなわち、少なくとも1つの後証された文仮説) が生 成される。文仮説検証器14は、意味情報25、辞書2 3、部分艦モデル26および反部分艦モデル27を使用 1 によって提供され意味仮脱生成器 1 3 によって使用さ れるため、検証された文仮脱け本質的に、直接に対応す する。キー旬に付けられた意味タグが、キー旬検出器1 る「意味」を有し、それにより、個々のアプリケーショ 的かつ意味的に最良の意味仮脱が検証され、最終出力 ンによる必要に応じた意味フレームの生成が可能とな

または半自動的に(すなわち、人間の支援のもとで)生 句検出を実行する。これは、会話状態に依存する特定の キー句検出器11によってキー句文法21から選択され **【0023】 [キー旬検出] キー旬検出器11は、キー** サブタスクに基づくことが可能である。具体的には、各 サブタスクごとに、キー句パターンが1つ以上の決定性 る。これらの文法は、タスク仕様から直接に人手により 単出することも可能であり、あるいは、当策者に関知の 従来の学習手続きを用いて、小さいコーパスから自動的 有限状態文法として記述される。これは、奥施例では、 成することも可能である。

込まれたキー句を含む句パターンを形成するために使用 [0024] 一般に、キー句は、従来のキーワードに加 り安定なマッチングが可能となり、検出精度が改善され (例えば、前掲のJ. Kawahara et al., "Concept-B ased Phrase Spotting Approach for Spontaneous Spee えて、"at the"や"near"のような機能語を含む。これに より、従来のキーワードのみのマッチングに比べて、よ れないがしばしばキー句に伴う充填句も定義され、埋め ch Understanding"を参照。) いずれのキー何にも含ま

て、キー句は繰り返し現れ、ガーベジモデルがキー句の 出現の間に埋め込まれる。しかし、注意すべき点である が、単純な繰り返しは曖昧さを生じろ可能性がある。例 【0025】特に、キー旬および充填句の文法はネット ワークへとコンパイルされる。このネットワークにおい て、不可能なキー句の結合を禁止する追加の砌約も組み えば、日の繰り返しが許容される場合、"twenty four" と"twenty"+"four"を区別することはできない。 従っ

【0026】従って、検出ユニットは、許容される結合 および反復を有するキー句部分文法オートマトンのネッ トワークからなる。このようなオートマトンは、結合塩 みを評価することによって、確率的質語モデルへと容易 に拡張することができる。このようなモデルを使用する ことにより、文レベルの文法と比べてあまり複雑になら ずに、適用範囲が広くなる。

「データ取得」サブタスクに適用された場合に、図1の 実施例のシステムのキー句検出器 1 1 によって使用され ることが可能である。このネットワーク例の完全な実現 により、曜日、月、日、および年の奥質的に任意の反復 が、適当な制約のもとに許容される。(このような完全 な実現の全語彙は99話である。) この特定のサブタス [0027] 例として、図2に、単純化した(すなわ ち、簡略化した)旬ネットワークの例を示す。これは、 クでは、キャリア句は組み込まれない。 20

[0028]さらに具体的には、ここに記載する本発明 の実施例によって採用されている検出方法は、フォワー ドーバックワード2パス探索に基づくものである。これ Resolution and Efficient Multiple String Hypothesi zation for Continuous Spaech Recognition Using Int er-Word Models", Proc. IEEE-1CASSP, volume 2, pp. 1 53-156, 1994、に記棋されている。 本発明の別の実施例 では、代わりに、当業者に周知の1パス被出社を使用す は、例えば、W. Chou et al., "An Algorithm of High ることも可能である。

換わった類似の単語列である。本発明の目標は、(入力 I. Kawahara et al., "Concept-Based Phrase Spotting 記載されているもの)は、N番目までの最良ストリング 仮脱からなる集合を求めることができるが、この結果と して得られるN層の敬良仮説は一般に、1~2個が置き 発話全体に基づいてストリング仮殻を生成することでは なく)入力発格の一部に基づいてキー句候補を疑別する ことであるので、仮説を延貰しても既に延長された仮説 [0029] A・認容スタック復号器 (例えば、前掲の Approach for Spontaneous Speech Understanding"¦ と同じ仮説になる場合にはその仮説は捨てられる。

40

付けることによって実現される。当業者には周知のよう [0030] 特に、本実値例のスタック復号器は、キー 句ネットワークのマージング (merging) 状態にマークを

20

し、さらに延長すると次の(すなわち新たな)句の最初 に、マージング状態は、キー句あるいは充填句が終了 に侵入することになるノードに対応する。

いている場合、本発明の手続きは、もう1話だけその句 検出した句を出力した後に現在の仮説は捨てられる。そ うでなければ、その時点は、その後の探索のためにマー [0031] スタック復号器によって「ポップ」される 仮説に、出力されるべき完全な句であるというタグが付 を延長し、その句を敬良延長と並べる。このノードに、 以前のいずれかの仮説が同じ時点に到達している場合、 クされる。

01

は、所留の個数の句を生成したことに基づいて、あるい は、あるスコアしきい値に基づいて、終了することも可 能である。例えば、仮説のスコアが、最高スコア仮説の 0. 99倍より小さい値に到達したときに、検出を終了 は、冗長な仮脱延長のない効率的なものであり、スコア 順に、正しいN番目までの最良のキー句族補を生成す る。本発明のさまざまな実施例によれば、この手続き [0032] 注意すべき点であるが、この検出手機き することも可能である。

は、次式のような従来の尤度比(LR(likelihood rati [0033] [キー句検証と信頼性尺度] 図1の実施例 のシステムのキー句検証器12は、部分語レベルのデス トに基ろいて、検出された句の検証を行う。具体的に は、与えられた句の各部分語nに対して、検証スコア 0) テストに基づいて計算される。

1 og LRn= (1 og P (O | 1,0°) - 1 og P (O | 1,0°)) / 10

そのものであるので、上記の計算の効果は単に、計算さ れるスコアを反部分間モデルのスコアだけずらし、その 注意すべき点であるが、式(2)の第1項は認識スコア (2) 結果を正規化することである。

ごとに、対応する部分語レベルの検証スコアを組み合わ※ CM= f (1 og LR1, [0036]キー旬検証器12は、検出された各キー句

40 ば、第1の例示的な信頼性尺度CMiは、フレーム継続 ★ 信頼性尺度 (CM) が、ある所定のしきい値を超える場 合に、与えられたキー句は承認される。実施例では、し 【0031】本発明のさまざまな実施例において、さま ざまな信頼性尺度関数を使用することができる。例え きい値の値は、例えばーロ、15に設定される。

 $CM_1 = \frac{1}{L} \sum_n (l_n \cdot \log LR_n)$

30 えられたキー旬のすべての部分語の対数尤度比の単なる 平均である。 (一実施倒では、句セグメンテーション後 話セグメントによる正規化に基づく。特に、これは、与 上記の式で、1nは、部分語nの維続時間を表し、しは [0038] 第2の例示的な信頼性尺度CM1は、部分 町の全継機時間である。すなわち、1−21,1である。

特開平10-116094 8

が得られる。(ビタビアルゴリズムは、当業者に周知の モデル22から得られ、対応する反部分語モデルは反部 分語モデル24から得られる。) 認識の結果として、観 および反部分語モデルを装す。 (部分語モデルは部分語 密列Oは、部分語nに対して、ピタピアルゴリズムによ ただし、Oは、観測フレームの列を扱し、スパおよびス ****は、それぞれ、部分語ヵに対する正しい部分語モデル** り並べられ、確率P (0 | ス nº) およびP (0 | ス nº) *LRn=P (0 | 10°) /P (0 | 10°) 従来のスコアリング方法である。)

ってなされる置換誤りを拒絶する能力が増大する。この (検証) ステップでは、文脈独立な反部分語モデルを使 (クラスタ化) によって構成される。各反即分語モデル 幽数の状態およびミクスチャを有する。 反部分語モデル り、部分語モデルの無制約な彼号を行うのに比べて、弁 別性が改善される。こうして、システムは、認職器によ 用することも可能であるが、認職ステップは、文脈依存 [0034] 各部分語モデルに対して、対応する反部分 は、対応する部分語モデルと同じ構造、すなわち、同じ は、特定の部分語の検証専用であるため、反部分語モデ 話モデルは、谯同しやすい部分語クラスをまとめること ルをリファレンスとして使用して復号を行うことによ

[0035] 特に、上記の式 (1) の対数をとり、その 結果を、観測Oの継続時間長1mに基づいて正規化する ことにより、量10g LRnが次のように定義される。 の部分語モデルを用いて実行される。

30 ure))を計算する。例えば、検出されたキー句がN個の ※せることによって、信頼性尺度(CM(confidence meas 部分語を含む場合、このキー句に対する信頼性尺度は、 対応するN個の尤度比の関数とすることが可能である。 具体的には、次のようになる。

★時間による正規化に基づく。特に、これは、正しい部分 **語モデルに対して得られるビタビスコアと、対応する反** 部分語モデルに対して得られるビタビスコアの差に等し い、すなわち、みのようになる。 (3) og L.R.

€

に単語間文脈情報が失われるため、最後の部分語に対し て特別の考慮がなされる。)すなわち、次のようにな

[数2]

3 Ξ $CM_2=rac{1}{N}\sum_n \log LR_n$

度)を受け取ることになる。従って、これを確実に拒絶* 句"December"の後半と完全に一致するため、部分語スコ アを平均した場合に高い検証スコア(すなわち信頼性尺 語は入力句とは非常に異なることがあるからである。例 ての部分語にわたる平均の信頼性レベルではなく、検証 プロセスの結果、信頼性レベルが低いような部分語に注 目する。これが有効なのは、正しくないキー句の部分語 のうちには実際に入力句に正しく一致するが、他の即分 【0039】類3の例示的な信頼性尺度CMがは、すべ えば、"November"の後半部分は、場合によって、入力

$$CM_3=rac{1}{N_s}\sum_{eta}ig\{egin{array}{ccc} \log LR_n & \log LR_n < 0 の場合 \ 0 & ext{ $ext{ }} \end{array}$$$

より小さい部分語の数 (すなわち、10g LRn<0と ただし、Naは、対数尤度比が実際には期待される平均 なる部分語の数) である。

【0041】第4の例示的な信頼性尺度CM4はシグモ ※20

$$CM_{\star} = \frac{1}{N} \sum_{n} \frac{1}{1 + \exp\left(-\alpha \cdot \log LR_{n}\right)}$$

これらの信頼性尺度のそれぞれに対して(あるいは、本 て)、特定のしきい値を選択することが可能である。 与 発明の別の実施例によって使用される信頼性尺度に対し えられた偖類性尺度の値がそのしきい値を下回る場合、 候補キー句は検証済みキー句候補の集合から排除され、 そうでない場合、検証済みキー句候補の集合に含まれ

「再スコアリング」を行うための基礎としても使用可能 である。例えば、E. Lleida and R. C. Rose, "Efficie nt Decoding and Training Procedures for UtteranceV erification in Continuous Speech Recognition", Pro c. IEEE-ICASSP, pp.507-510, 1996、には、北奥比に基 **ろいて復号を行うことが提案されている。しかし、尤度** 比の直接の使用は、そのダイナミックレンジが大きいた め、不安定となる可能性がある。こうして、本発明の一 実施例によれば、反部分語モデルのスコアが正しい部分 ベジ充填句は、もとの句と同じ継続時間を有し、もとの 果、もとの句は、その後の文解析(以下参照)で選択さ [0042] 本発明のさまざまな実施例によれば、計算 盾モデルのスコアより大きい場合(すなわち、CMiく 反部分語モデルをガーベジモデルとして処理する。ガー 旬よりも例えばCMIだけ高いスコアを有する。その結 される信頼性尺度の尤度比は、「フォールスアラーム」 0の場合)にガーベジ充壌句を生成することによって、 を排除するためだけではなく、検証済みの句に対する

【0043】 [文解析] 図1の実施例のシステムの文仮 **れる可能性が低くなる。**

特開平10-116094

8

*するためには、この句の前半(その検証スコアは低くな る可能性が高い)に注目するのが有効である。

び分散を計算する。その後、対数尤度比が、期待される 【0040】このように、低い信頼性レベルの部分語に 注目するために、各部分語ごとに正規分布を仮定するこ 具体的には、部分語HMMのトレーニングで用いたサン ブルを使用して、各部分語ごとに対数尤度比の平均およ 平均より小さい部分語のみを含む和を実行することによ って、CM1を計算することができる。すなわち、次の とによって、対数尤度比を調整することが可能である。

ようになる。 [数3]

9

9

※イド閲数を用いる。この例示的な信頼性尺度は、吸小鹍 り容易準でトレーニングするための損失関数として用い

3 [数4]

脱生成器 1 3 はキー句検証器 1 2 によって生成された検 られる。すなわち、次のようになる。

証済みキー句候補を意味情報25に甚づいて1個以上の 文仮説へと組み合わせる文解析を実行する。一実施例で である。ラティス解析法は、音響スコアと、提供される 意味制約情報 (キー句タグの許容される組み合わせを指 定する)に基づいて、句候補を結合する。キー句検出の ためのフォワードーバックワード探索によって与えられ 【0044】最も可能性の高い文仮説を効率的に見つけ **るためには、スタック復号探索法を採用すると有効であ** る。この方法は、一連の部分仮説を反復的に生成し、完 金な文仮脱が生成されるまで、各反復において最良の利 れる。別の実施倒では、島駅動保奈アルゴリズムを用い これは、トレリス解析よりわずかに精度が低くなるだけ は、句候補のLR(left-to-right)トレリスを使用する ことが可能な1次元RL(right-to-left)探索が用いら め、さらに別の実施例ではラティス解析法を採用する。 5スコアを音響スコアとして用いることが可能である。 ることも可能である。トレリス解析は計算量が多いた 用可能な部分仮説を延長する。

[0045] 具体的には、現在の「最良の」部分仮設を 対する上服スコアからのずれ (オフセット) として以下 (wi, wi) とし、新たな仮説が句wiを連結すること wz, wa} に対する評価関数は、完全な入力発話hoに によって生成されると仮定する。新たな仮説 (wi, のように計算される。

文仮説の意味的「完全性」を評価する。例えば、本発明

ただし、f '(wi)は、検出された句wiに対する評価 る。上限16は、認識プロセスのフォワードパスで計算 新たな句が追加されるごとに、オフセットが核算され の結果である。初期仮説は「「(null) = hoである。

50 01 は、本発明の一実施例によれば、スキップ長に比例する 紫となる可能性がある。従って、これを補うために、で timal Search Strategies for Speech Understanding C ることを仮定しないため、数語の短い仮説が誤って受容 される可能性が高い。従って、発話でスキップされた部 一様な罰金値をオフセットとして追加することが可能で ある。もちろん、この近似は相雑であるため、改善の探 きるだけ多くのキー句とともに、できるだけ多くのガー ベジ句 (無音を含む)を生成することが好ましい。 (一 実施例では、これらの仮説は、キー句検証プロセス中に 2、に記載されているような不足法(short-fall method) に基づいている。注意すべき点であるが、この評価はA 検出ベースの解析段階では特に、入力発話全体が扱われ ontrol", Artificial Intelligence, 18:295-326, 198 認容である。しかし、探索を効率的に導くこの方法の 発見的能力はやや限定されたものとなる可能性がある。 分を評価することが有効となる。そのため、具体的に [0046]上記の方法は、例えば、W. A. Woods, 生成することも可能である。)

は、大域的音響情報および大域的意味情報の両方を使用 【0047】 [文検証] 図1の実施例のシステムの文仮 説検証器14は、認識出力の最終判定を行う。 実施例で し、それぞれ入力発話全体に適用される。キー句検証ブ ロセスは局所的な判定のみをしたが、文仮殻検証プロセ スはこれらの局所的な結果を組み合わせ、従来の発話検 訊と同様の効果を実現する。しかし、検出ベースの認識 プロセスは、多数の予期しないキャリア句を含む場合で も入力発話を受容することが多いことに注意すべきであ

行される音響検証プロセスは、与えられた文仮説が十分 行われる。この段階で適用される部分語モデル (すなわ モデル24)よりも精度が高い。こうして、より高い精 [0048] 具体的には、文仮院検証器14によって実 に一致することを保証するために、入力発話全体の再ス ル26、反朝分哲モデル27、および辞書23を用いて ち、部分踏モデル26)の集合および対応する反部分語 モデル(すなわち、反部分語モデル27)の集合は、キ 一句検出器11およびキー句検証器12によって使用さ れるもの(すなわち、部分踏モデル22および反部分睛 コアリングを行う。この再スコアリングは、即分離モデ 度の音響再スコアリングが実行される。

50 [0049] 一方、意味検証プロセスは、与えられた各

段に基づいて実行される。このような場合、文仮説検証 ば、ユーザはただ"August" (8月) と言うだけで、その 月の特定の日を指定しないことがある。一般に、こうし [0050] 従って、本発明の一実施例によれば、文仮 脱検証器14は、与えられた文仮脱が意味要現を完成し この組合せ「テスト」は、例えば、満足な文仮説に遭遇 の一実施例によれば、意味検証は、ある構成要素が意味 的に「合法」かどうかのみを指定する単純な意味制約情 器14の意味解析部分は、例えば、与えられた文仮説の 意味要現が完全であるかどうかを判断する。しかし、注 意すべき点であるが、会話ベースのアプリケーションで ておらず、かつ、ほとんどの入力セグメントが弌寅比テ ストで拒絶された場合にのみ、その文仮散を拒絶する。 は、例えば、不完全な発話にしばしば遭遇する。例え た「不完全な」発話も同様に受容すべきである。

て、音響スコアとともに意味スコアを求め、組み合わせ - 般的な確率的意味モデルを、文仮説検証器14で用い [0051]しかし、本発明の別の実施倒では、さらに ることが可能である。このような場合、各文仮説につい たスコアを用いて、最終認識結果として出力すべき検証 するまで、各文仮説に適用することが可能である。

的意味モデルを用いている場合には、検証済み文仮説が 証のみまたは音響的検証のみ (両方ではない) を、文仮 らに高い精度の部分語および反部分語のモデルが利用可 能でない場合には、入力発話の音響再スコアリングを実 珠検証のみを実行して、単に、与えられた文仮貌が意味 **数現を完成していることを検証するか、あるいは、確率 敷終認証結果として判断されるもとになる意味スコアを** [0052] 本発明のさらに別の実施例では、意味的検 **敷検証器14で実行することが可能である。例えば、さ** 行することはあまり効果がない。従って、この場合、恵 された文仮睨を選択することが可能である。

30

[0053] [付記] 説明を明確にするため、ここに記 載した本発明の実施例は、園別の機能ブロックからなる ものとして殺した。これらのブロックによって扱される 機能は、共用あるいは専用のハードウェアの使用によっ トウェアを実行することが可能なハードウェアが含まれ るが、これに限定されるものではない。例えば、ここに 記載した構成要素の機能は、単一の共用プロセッサによ って、あるいは、複数のプロセッサによって提供するこ とが可能である。本発明の変施例は、ディジタル信号ブ ロセッサ (DSP) ハードウェア、上記の動作を実行す て堪供することが可能である。ハードウェアには、ソフ M)、および、結果を格納するランダムアクセスメモリ るソフトウェアを格納する謎み出し専用メモリ(RO 40

Ξ

(10)

[図2] 「日付取得」サブタスクに適用した場合に、図 1の例示的なシステムによって使用されることが可能な 単純化された句ネットワーク例の図である。

(R AM) からなることが可能である。超大規模集積 (V L S I) ハードウェアや、カスタムV L S I 回路を 孔用プロセッサやD S P回路と組み合わせたものも可能

[符号の説明]

1.1 キー旬検出器 文仮說生成器 キー句検証器 1 2

【0054】また、「キー句検出器」、「キー句検証器」、「文仮投身成器」、および「文仮投資証器」という用語は、対応する機能を実行する任意のメカニズムを

文仮說検証器 キー句文法 1 3 14

部分語モデル 钟选 2

【発明の効果】以上述べたごとく、本発明によれば、固

[0055]

定した形式的な文法に基づいて受容されるよりも多くの 発話が受容される、効率および自由度の高い会話音声器

反部分語モデル 53

3NO

8

TWENTY

TWENTY

3M0

反部分語モデル 部分語モデル 26

[図1] 本発明の実施例による音声認識および音声発話

[図面の簡単な説明] 職が実現される。

の理解を実行するシステムの図である。

[図]

[图2]

(13)

C-END

01

加动文 ローキ式れる (ササセをブレチ) 出跡 ムーママ神意と加利文式パち頭的 217 71-115 器出錄 器弧혀 器观主 烷 器 疏 教 笳承文 伊一丰 ローキ 57-LZ-33 諸代語図 **パモ**チ 翻代階 パモチ 部代語 バモチ 部公昭 れてチ 积盘 辟衛 女徒 52 4-4

各铭

3N0

THE C-BECIN

LUKZI

C-8ECIN

FIRST

MINETY

YADMUZ

FOUR

6

C-EECIM

MINETEEN

JANUARL

フロントページの様き

(71) 出版人 596077259

Murray Hill, New Jersey 07974-0636U, S. A. 600 Mountain Avenue,

(72)発明者 ピンーホワン ジャン

アメリカ合衆国、07059 ニュージャージ 一、ウォレン、サウス レーン 8

<u>|</u> |-

-10-

₹

(15)

特閒平10-116094

(12)発明者 チンーフイ リー アメリカ合衆国、01974 ニュージャージ - 、ニュー プロビデンス、ラニーメデ パークウェイ 118

(72)発明者 かわはら たつや 京都府京都市伏見区東泰行伏見御堂122

. . .

NETWORK INTERACTIVE USER INTERFACE USING SPEECH RECOGNITION AND NATURAL LANGUAGE PROCESSING

Publication number: JP2002524806T

Publication date:

2002-08-06

Inventor: Applicant: Classification:

- international:

G06F3/16; G06F13/00; G06F17/27; G06F17/28; G10L15/18; G10L15/26; H04L29/06; G06F3/16; G06F13/00; G06F17/27; G06F17/28; G10L15/00;

H04L29/06; (IPC1-7): G06F17/28; G06F3/16; G06F13/00

- European:

G06F3/16; G06F17/27; G10L15/18C2; G10L15/26A; H04L29/06

Application number: JP20000569391T 19990908

Priority number(s): US19980150459 19980909; US19980166198 19981005;

WO1999US20447 19990908

Also published as:

CA2343150 (A1)

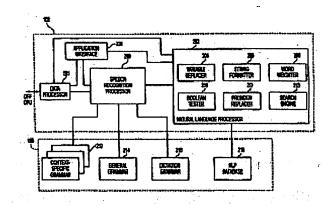
WO0014728 (A1) EP1110206 (A1) US6434524 (B1) EP1110206 (A0)

more >>

Report a data error here

Abstract not available for JP2002524806T Abstract of corresponding document: WO0014728

A system and method for interacting with networked objects, via a computer using utterances, speech processing and natural language processing. A Data Definition File relates networked objects and a speech processor. The Data Definition File encompasses a memory structure relating the networked objects, including grammar files and a natural language processor. The speech processor searches a first grammar file for a matching phrase for the utterance, and for searching a second grammar file for the matching phrase if the matching phrase is not found in the first grammar file. The system also includes a natural language processor for searching a data base for a matching entry for the matching phrase; and an application interface for performing an action associated with the matching entry if the matching entry is found in the database. The system utilizes context-specific grammars, thereby enhancing speech recognition and natural language processing efficiency. Additionally, the system adaptively and interactively "learns" words and phrases, and their associated meanings.



Data supplied from the esp@cenet database - Worldwide

報(*) ধ 盐 华 表 ধ (15)

(19) 日本国格許庁 (JP)

1) teratore 1 Equivalent to Pair Art (11)特許出顧公表番号

特表2002-524806

(P2002—524806A)	十成14年8月6日(2002.8.6) デ-73-ド(参考)	(43) ZEZC H
(P2002-524806A)	(43)公费日 平成14年8月6日(2002.8.6)	(43)公农日
	(P2002-524806A)	

觀別配与

(51) Int.Cl.7

5B091				
>	340A	340N	5 5 0 A	
17/28	3/16		13/00	
G06F				
	340		550	
17/28	3/16		13/00	
G 0 6 F				

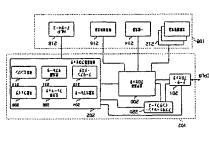
(全 49 頁) 枰 予備審查顧求 審査關水 未開水

(21) 出資番号 (86) (22) 出頭日 (85) 翻吹文提出日 (86) 国際山顧港号 (87) 国際公開日 (87) 国際公開日 (37) 國先推主張春号 (32) 優先日	(21) 出版部号 特額2000—568391(P2000—568391) (71) 出版 (85) (22) 出版日 平成11年9月8日(1998.9.8) (85) 翻形文推出日 平成11年3月9日(2001.3.9) (85) 国際山間部号 P C T Y U S 9 9 2 0 4 4 7 (87) 国際小間 P C T Y U S 9 9 2 0 4 4 7 (87) 国際小間日 P C T Y U S 9 9 2 0 4 4 7 (87) 国際小間日 P C T Y U S 9 9 2 0 4 4 7 (87) 国際小間日 P C T Y U S 9 10 10 10 10 10 10 10 10 10 10 10 10 10	(71) 出國人 5rature ((72) 発明者	(T)) 山圏人 ワン ボイス テクノロジーズ インコーボレーテッド アメリカ合衆国カリフォルニア州9212 サン ディエゴ スイート240 グリーン はUTE
(31)優先権主張番号	(31)優先権主張番号 09/166, 198	(74)代理人	(74)代理人 弁理士 恩田 博宜 (外1名)
(32) 優先日	平成10年10月5日(1998.10.5)	トターム(数	Fターム(勧考) 5B091 AA15 CA12 CA14 CB12 CC03
(33) 優先權主撰国	米国 (ns)		CD03

(54) 【発明の名称】 音声認識および自然音節処理を使用したネットワーク用対話型ユーザ・インタフェース

最終頁に嵌く

一タを介してネットワーク・オブジェクトと対話するシ 察し、もし上配第1文法ファイル内で上配照合対象語句 が発見されなければ上配照合対象語句に関して第2文法 発語、音声処理および自然言語処理を使用してコンピュ **ステムおよび方法。データ定義ファイルは、ネットワー** データ定義ファイルは、文法ファイルを含むネットワー けるメモリ構造を包含する。音声プロセッサは、上配発 **踊に対する服合対象節句に関して第1文法ファイルを検** ファイルを検索する。上配システムはまた、上配服合対 象節句に対する一致エントリに関してデータベースを検 **客する自然質語プロセッサ;および、もし上記一致エン** トリが上記データペース内で発見されたならば該一致エ ク・オブジェクトを自然言語プロセッサに対して関連付 ク・オブジェクトおよび音声プロセッサを関連付ける。



[特許請求の範囲]

[請求項1] ネットワーク・オブジェクトと関連付けられたネットワーク オブジェクト・テーブルをネットワークを介して転送する工程と、 上記ネットワーク・オブジェクトに対する一致エントリに関して上記ネットワ ーク・オブジェクト・テーブルを検索する工程と、 上記ネットワーク・オブジェクト・テーブル内において上記一致エントリが発 見されたなら上記一致エントリに関連付けられた動作を実施する工程と

を含む、ネットワーク・オブジェクトとの音声対話の為にコンピュータを更新 する方法 【請求項2】前記ネットワーク・オブジェクト・テーブルはウェブ・サイト から読み出される、請求項1記載の方法。 [請求項3] 前記ネットワーク・オブジェクト・テーブルは、複数のネット ワーク・オブジェクトに対して各ネットワーク・オブジェクト・テーブルを記憶 する箇所から読み出される、請求項1記載の方法。 【請求項4】前記ネットワーク・オブジェクト・テーブルは、文脈固有文法 も含むダイアログ定義ファイル内に含まれる、請求項1記載の方法、 【請求項5】前記ネットワーク・オブジェクト・テーブルは、自然言語プロ セッサ・データベースに対する各エントリも含むダイアログ定義ファイル内に含 まれる、請求項1記載の方法。 【請求項6】前記ネットワーク・オブジェクト・テーブルは、文脈固有文法 および自然言語プロセッサ・データベースに対する各エントリも含むダイアログ 定義ファイル内に含まれる、請求項1記載の方法。

ファイルであってネットワーク・オブジェクト・テーブルを含むダイアログ定義 【請求項7】ネットワーク・オブジェクトと関連付けられたダイアログ定義 ファイルを転送するネットワーク・インタフェースと

上記ネットワーク・オブジェクトに対する一致エントリに関して上記ネットワ **ーク・オブジェクト・テーブルを検索するデータ・プロセッサと**、 上記ネットワーク・オブジェクト・テーブル内において上記一致エントリが発 見されたなら上記一致エントリに関連付けられた動作を実施するアプリケーショ

の効率を高める。更に上記システムは、単語および語句

ならびにそれらの関連意味を適合的にかつ対話的に"学

ン・インタフェースも含む。上記システムは文脈固有文

法を利用することにより、音声認識および自然昏語処理

ントリに関連付けられた動作を実施するアプリケーショ

-2-

ソ・インタフェースと、

を含む、ネットワーク・オブジェクトとの音声対話の為にコンピュータを更新 するシステム。 【請求項8】ネットワーク・オブジェクトと関連付けられたダイアログ定義 ファイルであってネットワーク・オブジェクト・テーブルを含むダイアログ定義 ファイルを発見する工程と

上記ダイアログ定義ファイルを読み出す工程と、

上記ネットワーク・オブジェクトに対する一致エントリに関して上記ネットワ ーク・オブジェクト・テーブルを検索する工程と、 上記ネットワーク・オブジェクト・テーブル内において上記一致エントリが発 見されたなら上記一致エントリに関連付けられた動作を実施する工程と、 を含む、ネットワーク・オブジェクトとの音声対話の為にコンピュータを更新 する方法 【請求項9】前記ダイアログ定義ファイルはウェブ・サイトから読み出され る、請求項8記載の方法。 【請求項10】前記ダイアログ定義ファイルは複数のネットワーク・オブジ ェクトに対するダイアログ定義ファイルを記憶する箇所から読み出される、請求 項8記載の方法。

【請求項11】前記ダイアログ定義ファイルは記憶媒体から読み出される、 請求項8記載の方法。 [請求項12] 前記ダイアログ定義ファイルは文脈固有文法を含む、請求項 8 記載の方法。

[請求項13] 前記ダイアログ定義ファイルは自然言語プロセッサ・データ ベースに対するエントリを含む、請求項8記載の方法。 [請求項14] 前記ダイアログ定義ファイルは文脈固有文法と自然言語プロ セッサ・データベースに対するエントリとを含む、請求項8記載の方法。

発明の詳細な説明】

特扱2002-524806

€

[0001]

(発明の背景)

1. 発明の分野

1 9 98年9月10日に出願されて"音声認識および自然言語処理を使用した対話型 本発明はネットワーク・コンピュータ用ユーザ・インタフェースの音声認識に 関する。より詳細には本発明は、音声認識および自然言語処理を使用してユーザ ューザ・インタフェース"と称された米国特許出願第09/150,459号の 本発明は、 とコンピュータが対話する新規な方法およびシステムに関する。 一部継続出願である。

[0002]

11. 関連技術の説明

コンピュータが普及するにつれ、コンピュータの理解およびコンピュータとの **対話が多くの人々に取り困難であることが明らかとなっている。コンピュータ操** 作の為にユーザは、原始的なコマンドおよび直感的でない手順を学ぶべき場合が ンドウズ(登録商標)式のオペレーティング・システムを使用する。故にユーザ 多い。例えば殆どのパーソナルコンピュータは、概してメニュー主導であるウィ は、どのメニュー・コマンドもしくは一連のコマンドが所望結果を与えるのかを 学ぶ必要がある。

[0000]

更に、コンピュータに対する習用の対話はキーボードもしくはマウスなどの手 動入力デバイスの故に遅くなることも多い。多くのコンピュータ・ユーザは熟練 、ならびに、ウェブ・サイトなどのネットワーク・オブジェクトと対話する為の これらの手動入力デバイスによりコンピュータ コンピュータ にコマンドおよび単語を入力するには相当の時間が掛かる。故に、 容易かつ高速で更に直感的な方法が必要なことは明らかである。 タイピストではない。その結果、

[0004]

コンピュータとの対話の為に提案されたひとつの方法は、音声認識である。音 **声認識は、人間の音声を音響的に検出すると共に検出された音声を一連の単語へ** 9

特級2002-524806

と翻訳すべく恊働するソフトウェアおよびハードウェアを必要とする。当業界に ら成り;第1の音素は"th"音であり、第2の音素は"o"音であり、第3の おいて公知である如く音声認識は、ハードウェアが検出した音を、音繋(pho 音素は別個の音の単位である。例えば"those"という単語は3個の音素か 音素は"s"音である。音声認識ソフトウェアは、検出された各音素を記憶辞書 agon Systems, Inc. に譲渡された米国特許第4, 783, 80 3 号に与えられるが、該米国特許第は言及したことにより本明細書中に援用する 現在では、Dragon Systems, Inc. およびInternat ional Business Machines Inc. などの会社から市 998年11月8日に発行されて"音声認識装置および方法"と称され且つDェ からの既知の各単語と一致させるべく試行する。音声認識システムの一例は、 neme)と称される分割不能な更に小さな音へと分解することで機能する。 販された多くの音声認識ソフトウェア・パッケージが在る。

[0005]

これらの音声認識ソフトウェア・パッケージもしくはシステムのひとつの制限)機能のみを実施することである。故にユーザは依然として、コンピュータを操 は、それらが通常はコマンドおよび制御もしくは聴音書取(dictation 作する為にはコマンドの用語集を学ぶ必要がある。

[0000]

これらの音声認識システムに対して提案された改良は、検出された単語を自然 自然言語処理は一般的に、検出 された単語の文法的関係および相対的文脈を分析することにより各単語の概念的 s Machines Inc. に譲渡された米国特許第4,887,212号 は言及したことにより本明細書中に援用するが、該米国特許第は単語分離、形態 1989年12月12日に発行されると共に"自然言語テキストに対するパーサ (PARSER)"と称され且つInternational Busines 学的分析、辞書検索および文法分析を使用して入力単語ストリームを解析する方 話し手が意図している意味)を決定する必要がある。 言語処理システムを用いて処理することである。 (例えば、

[0007]

はその自然言語処理システムが正しい"意味"を決定することを期待してその語 ボードもしくはマウスなどの手動入力よりも強力なコンピュータの操作ツールが 識システムにより検出された単語の正しい"意味"を自然言語処理システムが決 **定し得ない可能性があることである。斯かる場合にユーザは通常、次回の試行で** 句を言い換えたり言い直したりする必要がある。これは、ユーザが文もしくはコ マンドの全体を言い直す必要があるときに相当の遅延に繋がるのは明らかである 従来システムの別の欠点は、音声認識に必要な処理時間が法外に長くなり得る ことである。これは主として、処理されるべき大量の情報と比較して処理リソー は検索される辞書ファイルのサイズの故に、発語(utterance)の認識 スの速度が有限だからである。例えば従来の多くの音声認識プログラムはおいて 音声認識とともに自然言語処理を使用すると、話された単語を使用して、キー 提供される。しかし乍ら従来の自然言語処理システムのひとつの欠点は、 に要する時間は長い。

[0008]

従来の音声認識/自然言語処理システムの更なる欠点は、それらが対話型で無 ブジェクトとの間には新たなリレーションシップが形成される。故に従来の音声 クトにより課された新たなリレーションシップに起因する状況には対処し得ない 。結果として、コンピュータとの会話形式対話は不可能である。ユーザは、完全 いので新たな状況に対処し得ないことである。未知のもしくは新たなネットワー ク・オブジェクトにコンピュータ・システムが遭遇したとき、コンピュータとオ 認識/自然言語処理システムは、それまで未知であったネットワーク・オブジェ な概念をコンピュータに伝達する必要がある。ユーザはまた断片的な文を話すこ とが出来ない、と言うのも、(以前の発語の意味に依存する)それらの断片的な 文の意味が失われるからである。

[0000]

必要なものは、上述の各欠点を回避する音声認識および自然言語処理を括用し た、コンピュータとの対話型ユーザ・インタフェースである。

(発明の概要)

-5

6

び一般文法は上記聴音書取文法よりも物理的に小さいからである。

特数2002-524806

本発明は、発語、音声処理および自然言語処理を使用してコンピュータと対話 する新規で優れたシステムおよび方法である。概略的に上記システムは音声プロ 第1文法ファイルを検索し、且つ、もし上記第1文法ファイル内で上記照合対象 語句が発見されなければ上記照合対象語句に関して第2文法ファイルを検索する セッサを含むが、該音声プロセッサは、上記発語に対する照合対象語句に関して 上記システムはまた、上記照合対象語句に対する一致エントリに関してデータ ベースを検索する自然言語プロセッサ、および、もし上記一致エントリが上記デ **ータベース内で発見されたならば該一致エントリに関連付けられた動作を実施す** るアプリケーション・インタフェースも含む。

[0010]

タベース内で発見されなければ上記データベース、上記第1文法ファイルおよび 好適実施例において上記自然言語プロセッサは、上記一致エントリが上記デー 上記第2文法ファイルの少なくともひとつを上記照合対象語句により更新する。

上記第1文法ファイルは文脈固有文法ファイルである。文脈固有文法ファイル とは、特定主題に相当に関連する単語および語句を含むファイルである。上記第 て解釈の不要な単語および語句を含むファイルである。すなわち、上記一般文法 ファイル内の単語および語句は一切の起源文脈に属さない。本発明では上記一般 ーザは更に会話的な形式でコンピュータと意志疎通でき、その場合に話された単 語は、もし上記文脈固有文法ファイル内で発見されれば、最も最近に話された主 2 文法ファイルは一般文法ファイルである。一般文法ファイルとは、文脈に鑑み 文法ファイルを検索する前に上記文脈固有文法ファイルを検索することから、 題に鑑みて解釈される。

[0012]

本発明の更なる側面において上記音声プロセッサは、もし上記照合対象語句が 集である。上記文脈固有文法および一般文法を最初に検索することにより音声認 上記一般文法ファイル内で発見されなければ上記照合対象語句に関して聴音書取 文法を検索する。上記聴音書取文法は、一般的な単語および語句の大規模な用語 職時間が相当に短縮されることが期待されるが、これは、上記文脈固有文法およ

上記データベースを検 上記照合対象語句の少なくとも一個の単語をワイルドカードで置換する上記自然 言語プロセッサ内の変数リプレーサにより達成され得る。上記語句内の("単語 変数"と称される)一定の単語をワイルドカードで置換することにより、上記デ ータベース内のエントリ数は相当に減少され得る。更に上記自然言語プロセッサ **内の代名詞サブスティテュータは上記照合対象語句内の代名詞を固有名称に置換** することにより、上記データベース内にユーザ特有事実が記憶されるのを許容す 索する前に上記照合対象語句の少なくとも一個の単語を置換する。 本発明の別の側面において上記自然言語プロセッサは、 [0013]

[0014]

記データベースを検索する前に個々の単語の相対有意性に従い上記照合対象語句 本発明の別の側面において文字列フォーマッタは上記データベースを検索する 内の個々の単語を加重する。これらの工程によれば上記データベースの更に高速 前に上記照合対象語句をテキスト・フォーマットする。また単語ウェイタは、 かつ正確な検索が許容される。

[0015]

記一致エントリ内に存在するか否かを決定する。このブーリアン・テストは、上 上記自然言語プロセッサ内の検索エンジンは、上記一致エントリに対する確信 値を生成する。また上記自然言語プロセッサは上記確信値をスレッショルド値と 比較する。ブーリアン・テスタは、上記照合対象語句からの必要個数の単語が上 記検索エンジンにより戻される結果を検証する役割を果たす

[0016]

上記照合対象語句からの必要個数の単語が上記一致エントリ内に存在しなけれ トリが上記発語の正しい解釈であるか否かを問い質す。上記一致エントリが上記 発語の正しい解釈でなければ、上記自然言語プロセッサはユーザに対し付加的情 ば、曖昧さを排除すべく上記自然言語プロセッサはユーザに対し、上記一致エン 報を問い質す。この付加的情報により、上記データベース、上記第1文法ファイ

ルおよび上記第2文法ファイルの少なくともひとつが更新される。この様にして本発明は付加的な発語の意味を適合的に"学習"することにより、上記ューザ・インタフェースの効率を高める。

[0017]

上記音声プロセッサは、引き続く発語に対する引き続く照合対象語句に関連付けられた文脈固有文法を有効化して検索する。これにより、最も適切な単語および語句が最初に検索されて音声認識時間を短縮することが確実となる。

[0018]

概略的に本発明は、ウェブ・ページなどのネットワーク・オブジェクトとの音 オブジェクトを上記音声対話システムに関連付けるネットワーク・オブジェクト ・オブジェクト・テーブルを記憶する統合箇所にて、上記ネットワーク・オブジ ェクト内に埋め込まれる。上記ネットワーク・オブジェクト・テーブルは、上記 トワーク・オブジェクトにエントリが一致すると、テキスト音声がスピーカから 音声発生され、文脈固有文法ファイルが使用され、または、自然言語プロセッサ ジェクト・テーブルは、ダイアログ定義ファイルの一部とされ得る。ダイアログ サイトにて、または、複数のネットワーク・オブジェクトに対するネットワーク ネットワーク・オブジェクトに一致するエントリに関して検索される。上記ネッ ・データベースが使用されるなどの動作が行われ得る。上記ネットワーク・オブ 自然言語プロセッサ・デー ・テーブルがネットワークを介して上記コンピュータへと転送される。上記ネ 声対話の為にコンピュータを更新する方法を含む。最初に、上記ネットワーク トワーク・オブジェクト・テーブルの場所は、特定のインターネット・ウェブ タベースに対するエントリを、または両者を含み得る。 または、 定義ファイルは、文脈固有文法ファイルを、

[0019]

本発明の別の側面において、ネットワーク・インタフェースは上記ネットワーク上からダイアログ定義ファイルを転送する。上記ダイアログ定義ファイルはネットワーク・オブジェクト・テーブルを含む。データ・プロセッサは、上記ネットワーク・オブジェクトに一致するデーブル・エントリに関して上記ネットワーク・オブジェクト・テーブルを検索する。この一致テーブル・エントリが発見さ

れたなら、アプリケーション・インタフェースは上記一致エントリにより特定さ

特数2002-524806

9

[0020]

れた動作を実施する。

本発明の別の側面においては、ネットワークと関連付けられたダイアログ定義ファイルが発見されて読み出される。上記ダイアログ定義ファイルは、ウェブ・サイト、記憶媒体などの種々の場所から、または、複数のネットワーク・オブジェクトに対するダイアログ定義ファイルを記憶する場所から読み出され得る。上記ダイアログ定義ファイル内に含まれたネットワーク・オブジェクト・テーブルが検索される。上記一致エントリは上記ネットワーク・オブジェクト・デーブルが検索される。上記一致エントリは上記ネットワーク・オブジェクトと関連付けられた動作を定義し、その動作はその後に上記システムにより実施される。ネットワーク・オブジェクト・デーブルの他にも、上記ダイアログ定義ファイルは文脈固有文法を、自然言語プロセッサ・データベースに対するエントリを、または両者を含み得る。

[0021]

本発明の特徴、目的および利点は添付図面に関して以下に示された詳細な説明 から更に明らかとなるが、同一の参照符号は各図を通して対応している。

(好適実施例の詳細な説明)

次に、図1の好適なコンピュータ・システム100の機能プロック図を参照して本発明を開示する。図1においてコンピュータ・システム100は、中央処理コニット (CPU) 102を含んでいる。CPU 102は、図3A乃至図3Dに関して本明細書中で記述される方法を実施すべく適切にプログラムされた当業界では公知の任意の汎用マイクロプロセッサもしくはマイクロコントローラとされ得る。上記CPUをプログラムするソフトウェアは記憶媒体108から、又は、コンピュータ・ネットワークを介した別の場所から読み取られ得る。例えばCPU 102は、Intel社などにより製造されたPentium 11プロセッサなどの習用のマイクロプロセッサとされ得る。

[0022]

CPU 102は、ディスプレイ104、手動入力器106、記憶媒体108

特数2002-524806

マイクロホン110、スピーカ112、データ入力ポート114およびネット 録商標)・ディスク・ドライブ、CD-ROM、シリコン・メモリ、または、デ 一タの記憶および読出しの為に当業界で公知の他のメモリ・デバイスとされ得る 104は、ユーザに対してイメージおよびテキストを視覚表示する当業界で公知 のCRT、LCD画面、タッチパネルまたは他のモニタなどの視覚的ディスプレ データの手動入力の為に当業界で公知の他の入力デバ 重要な点として、記憶媒体108はCPU 102から離間して配置され得る ネットワーク (WAN) もしくはインターネットなどのネットワークを介してC PU 102に接続され得る。マイクロホン110は、CPU 102に対して オーディオ信号を提供する当業界で公知の任意の適切なマイクロホンとされ得る 。スピーカ112は、CPU 102からのオーディオ信号を再生する当業界で 12は、適切ならばデジタル/アナログおよびアナログ/デジタル変換回路を含 と共に、ローカル・エリア・ネットワーク(LAN)もしくはワイド・エリア・ 公知の任意の適切なスピーカとされ得る。マイクロホン110およびスピーカ1 バスなどのデータ・プロトコルを使用して外部付属品とインタフェースする当業 16は、コンピュータ・ネットワークを介してファイルを通信もしくは転送する 当業界で公知の任意のインタフェースとされ得るが、斯かるネットワークの例と インタフェース116はデータ入力ポート114に接続されたモデムから成り得 ワーク・インタフェース116などの複数の周辺装置と通信する。 ディスプレイ 界で公知の任意のデータ・ポートとされ得る。ネットワーク・インタフェース1 イとされ得る。手動入力器106は、習用のキーボード、ギーパッド、マウス、 み得る。データ入力ポート114は、RS-232、ユニバーサル・シリアル しては、TCP/IP、イーサネット(登録商標)、またはトークン・リング ネットワークが挙げられる。これに加え、一定のシステム上にてネットワーク イスとされ得る。記憶媒体108は、磁気ディスク・ドライブ、フロッピー トラックボール、または、

[0023]

故に図1は、コンピュータ・システム100の各機能要素を示している。コン ピュータ・システム100の各要素は、上述の如き適切な既製の構成要素とされ

得る。本発明は、音声を使用してコンピュータ・システム100との人的対話を 行う方法およびシステムを提供する。

[0024]

図8に示された如くコンピュータ・システム100は、コンピュータ・ネット ワークの集合であるインターネット700に接続され得る。インターネット70 0に接続する為に、コンピュータ・システム100はデータ入力ポート114に 接続されたモデムであるネットワーク・インタフェース116、または、当業界 で公知の他の任意の方法を使用し得る。ウェブ・サイト710、他のコンピュー タ120および記憶媒体108もまた、当業界で公知の斯かる方法によりインタ ーネットへと接続され得る。

[0025]

職プロセッサ200、データ・プロセッサ201、自然言語プロセッサ202お 次に図2を参照すると、図2はCPU 102および記憶媒体108の拡大機 **能ブロック図を示している。図2の各機能要素は全体的にソフトウェアもしくは** ハードウェアまたは両者にて具現され得ることは理解される。ソフトウェア実施 例の場合、そのソフトウェアは記憶媒体108にて、または、コンピュータ・ネ データ・プロセッサ20 ンタフェースする。データ・プロセッサ201によれば上記CPUは、これらの 1 および検索エンジン213を含んでいる。記憶媒体108は、複数の文脈固有 文法ファイル212、一般文法ファイル214、聴音書取文法216、および自 **炊言語処理(NLP)データベース218とを含んでいる。好適実施例において** 各文法ファイル212、214および216は、ユーザが話す言語の構造を記述 BNFファイルは言語の ディスプレイ104、記憶媒体108、マイクロホン110、スピーカ1 weighter) 208、ブーリアン・テスタ210、代名詞リプレーサ21 データ入力ポート114およびネットワーク・インタフェース116とイ ソースからデータを発見して読み出し得る。自然言語プロセッサ202は更に、 変数リプレーサ204、文字列フォーマッタ206、単語ウエイタ(word ットワークを介した代替的箇所にて読み取られ得る。CPU 102は、 するバッカス・ナウア記法(BNF)ファイルである。 よびアプリケーション・インタフェース220を含む。 1 13,

構造を記述するものとして当業界で公知であることから、BNFファイルの詳細 は本明細書中では論じない。 BNFファイルのひとつの利点は、語句もしくは単 語のシーケンスを記述する上でこれらの単語シーケンスの全ての組合せを明示的 に記述することなく、階層ツリー状構造が使用され得ることである。故に、該好 アイル212、214および216の物理的サイズは最小化され、以下に記述さ れる如くこれらのファイルが有効化されて検索され得る速度が高まる。但し代替 **適実施例においてBNFファイルを使用することにより記憶媒体108内の各フ** 実施例においては、他のファイル構造が使用される。

次に図3 A 乃至図3 D のフローチャートを参照し、図2 の各機能要素の動作お 、音声プロセッサ200に対して発語(utterance)を提供する。発語 とは、始まりおよび終わりを有する一連の音であり、話された単語を一個以上含 wav"オーディオ・ファイル・フォーマットなどのデジタル・フォーマットで よび相互作用を記述する。図3Aにおいて、フローはブロック300にて開始し み得る。話された単語を捕捉するマイクロホン110はブロック300の工程を 実施し得る。代替的に、発語はデータ入力ポート114を介して記憶媒体108 から音声プロセッサ200~と提供され得る。好適には、上記発語は公知の

[0027]

フローは判断302に進み、其処で音声プロセッサ200は、文脈固有文法2 12のひとつが有効化されているか否かを決定する。もし文脈固有文法212が 育効化されていれば、ブロック304にては各文脈固有文法212が検索される 。好適実施例において各文脈固有文法212は、起顔文脈(parent co ther)"もしくは"株 (stocks)"などである。斯かる場合に各文脈 実施例において、文脈の例は"ニュース(news)"、または"気象(wea 修飾語もしくはパラメータを含むことになる。各文脈を使用すると、情報の形式 文脈が対象領域である。例えばパーソナルコンピュータに適用可能な本発明の一 n t e x t)に関する単語および語句を含むBNFファイルである。概略的に、 固有文法212は、これらの文脈の夫々に対応するコマンド、制御単語、説明,

に対して階層構造が提供される。各文脈およびそれらの用法は、NLPデータベ 一ス218を参照して以下で更に説明する。

特表2002-524806

₹

[0028]

ク300にて提供された上記発語に対する一致に関して検索される。但し文脈固 有文法212が未だ有効化されていなければ、フローはブロック308へと進み もし文脈固有文法212が有効化されていれば、文脈固有文法212はブロッ 、其処で一般文法214が有効化される。

[0029]

好適実施例において一般文法214は単語および語句を含むBNFファイルで ル212が存在する関連文脈を有し得る。検言すると、一般文法214中の単語 **一夕に適用可能な一実施例において、一般文法214はコマンドおよび制御語句** あり、斯かる単語および語句自体は起源文脈に属しないが、文脈固有文法ファイ および語句は階層的文脈構造のルートに存在し得る。例えばパーソナルコンピュ を含むことになる。

[0030]

ブロック314にて聴音書取文法216が有効化される。好適実施例において聴 音書取文法216は各単語のリストを含む BNFファイルであり、各単語自体は 実施例において、聴音書取文法216は一般的辞書に類似した一般的単語の比較 ブロック310にて一般文法214は、ブロック300にて提供された発語に 起源文脈も関連文脈も有さない。例えばパーソナルコンピュータに適用可能な一 一致が見られたか否かに依存して判定が為される。もし一致が見られなければ、 対する照合対象単語もしくは語句に関して検索される。ブロック312にては、 的に大きなリストを含んでいる。

[0031]

ブロック316にて上記聴音書取文法は、ブロック300にて提供された発語 の各単語に対する照合対象単語に関して検索される。判定ブロック318にて照 合対象単語が発見されなければ、ブロック320にてはディスプレイ104に視 質的エラー・メッセージが選択的に表示され、または、スピーカ112により可 聴エラー・メッセージが選択的に再生される。而して上記プロセスは、ブロック

300にて音声プロセッサ200に対し別の発語が提供されるまで停止する。

[0032]

故に上記記述から理解され得る如く、音声プロセッサ200に対して発語が提 し一般文法214でも一致しなければ、聴音魯取文法316が有効化され且つ検 供されたとき、もし有効化された文脈固有文法ファイル212があれば、その文 脈固有文法ファイル212が先ず検索される。有効化された文脈固有文法ファイ ル212内で一致が無ければ、一般文法214が有効化され且つ検索される。 素される。

[0033]

戦文法214のいずれかを検索しているとき、それは"コマンド/制御"モード に在ると称される。このモードにおいて音声認識プロセッサ200は、発語全部 音書取モードにおいて一度に一単語ずつ検索することよりも一般的に高速である を全体として上記文法中の各エントリと比較する。これと対照的に、音声認識プ 一ドに在ると称される。このモードにおいて音声認識プロセッサ200は、聴音 好適実施例において音声認識プロセッサ200が文脈固有文法212または一 鲁取文法216内の各エントリに対し、上記発語を一度に一単語ずつ比較する。 ロセッサ200が上記聴音書取文法を検索しているとき、それは"聴音書取" 上記コマンド/制御モードにおいて発語全体に対する一致を検索することは、 と期待される。

[0034]

更に、任意の個々の文脈固有文法212は一般文法214よりもサイズが小さ く(すなわち合計の単語および語句が少なく)、一般文法214自体は聴音書取 文法216よりもサイズが小さいことも期待される。故に、有効化された任意の ル・サイズが小さい故に、一致があれば更に迅速に発見されると思われる。同様 に、聴音書取文法216の前に一般文法214を検索することにより、一致があ 文脈固有文法212を最初に検索すれば、少なくとも部分的な理由としてファイ れば更に迅速に発見されると思われる。

[0035]

また、文脈固有文法212および一般文法214の両者に対して適合的に付加

9

特数2002-524806

設的な発語を含んでいる。その故に、聴音書取文法216におけるよりも文脈固 有文法212もしくは一般文法214における方が更に迅速に一致が発見される される本発明の機能に関して以下で更に説明される如く、これらの文法は最も一 ことが期待される。

[0036]

音 うのも、それらの単語および語句はユーザが最も最近に話した主題に相当に関連 ューザの各単語の意味が解釈され乍ら、ユーザは断片的な文を使用して更に会話 最後に、以下で更に説明される如く、有効化された文脈固有文法212におけ する単語だからである。これにより、有効化された文脈固有文法212に従って る単語および語句はユーザにより発語される可能性が更に高いものである、 的な形式で話すことができる。

[0037]

び語句を含む単一の大きなリストで一度にひとつのエントリを検索するよりも更 本発明は上述のシーケンスで検索を行うことから、予測される全ての単語およ に効率的な検索を行い得る。

[0038]

判定306に戻ると、文脈固有文法212内で一致が発見されたなら、図3A に示された3つの可能的な次工程が在る。有効化された文脈固有文法212内の 各一致エントリに対し、音声認識プロセッサ200により行われる関連動作が在 り得る。ブロック322はひとつの動作がアプリケーション・インタフェース2 セスすること、または、テキストを音声に変換して特定語句を話すこと、が指示 20に指示を行い、該アプリケーション・インタフェース220が別体のソフト ウェア・アプリケーションもしくはエントリに関して一定の動作を行い得ること を示している。例えばアプリケーション・インタフェース220は、Micro s oftによる音声アプリケーション・プログラミング・インタフェース(SA P I)規格を使用して外部アプリケーションと通信し得る。上記外部アプリケー ンョンに対しては例えば、特定のインターネット・ウェブ・サイトURLにアク され得る。他の動作は、図4のNLPデータベース218に関して以下で更に論 じられる如く行われ得る。

[0039]

8 におけるひとつの行(row)に直接的にアクセスすることにより、以下で更 は語句に対し、有効化された文脈の単語もしくは語句を前方付属(prepen d) し得ることを示している。例えば有効化された文脈が"映画 (movies)"であり且つ照合対象発語が"8時(8 o'clock)"であれば、単語 326は別の動作として、文脈固有文法306で発見された照合対象単語もしく ブロック324は別の動作として、自然言語処理(NLP)データベース2] "映画"が語句"8時"に前方付属されて"8時の映画 (movies に記述される各自然言語処理工程をバイパスし得ることを示している。

[0040]

8 o'clock)"という語句が形成される。

同様に、もし一般文法214内に一致が発見されればフローは、ブロック32 2 に進み其処でアプリケーション・インタフェース220に上記の動作を行うこ とを指示し、または、ブロック324に進み其処でNLPデータベース内のひと つの行が直接的にアクセスされる。但し一般文法214において一致が発見され たときに文脈の前方付属は生じない、と言うのも、上述の如く一般文法214内 の各エントリはそれら自体が起源文脈を有さないからである。

[0041]

代替的に、ブロック301にては手動入力された単語が補捉され、上記自然言 語プロセッサヘと入力される。 最後に図3Aを参照すると、各単語は手動入力器106を介して手動で入力さ れ得る。この場合に音声認識は不要であるが、入力された単語の自然言語処理は 依然として望ましい。故にフローは図3Bへと進む

[0042]

図38のブロック328において、自然言語プロセッサ202は自然言語処理 ッタ206により達成されると共に、各単語間の二重空白を除去し、全ての文字 分析の為に上記語句をフォーマットする。このフォーマットは、文字列フォーマ t is"に変更する)などのテキスト処理を含み得る。このフォーマット工程 を小文字(もしくは大文字)にし、短縮を広げる(例えば、"it's"を

特級2002-524806

8

の目的は、語句を解析に対して準備することである。

[0043]

か (what movies are playing at time) " と 変換されるが、この場合に"時間 (time)"は任意の時間値を表すべく使用 d o l || IBM stock at dollars)"と変換されるが、この場 合に"ドル(dollars)"は任意のドル値を表すべく使用されるワイルド カード機能である。この工程は簡単なループにより達成され得るが、該ループは "ドル"もしくは"時 (o' c l o c k)" などの単語であるキー・トークンに 関して語句を検索し、且つ、特定のワイルドカード機能により単語変数を置換す るものである。置換が行われた語句内の箇所を追跡する為に配列が使用され得る これにより、NLPデータベース218が検索された後で一定の位置にて上記 フローはブロック330~と続き、該ブロックにては、NLPデータベース2 18へのアクセスに備えて変数リプレーサ204により関連ワイルドカード機能 により一定の"単語変数 (wordーvariable)" が置換される。本明 細書中で使用される"単語変数"とは、量、日付、時間、通貨などを表す単語も しくは語句を指している。例えば一実施例において"8時にはどの映画が放映さ 1 o c k) "という語句はブロック330にて"時間にはどの映画が放映される されるワイルドカード機能である。別の場合の一実施例においては、"IBMの lars)"という語句はブロック330にて"IBMの株をドルで売る (se , o 株を100ドルで売る (sell IBM stock at 100 œ れるか (what movies are playing at 語句へと元の単語変数を戻すべく再置換することが許容される。

[0044]

ブロック330にて単語変数を関連ワイルドカード機能により置換する目的は NL Pデータベース218内に存在すべきエントリの個数を減少することであ o, clock)、9邸(o' clock) などの別体のエントリ ではなく、"時間にはどの映画が放映されるか(what movies ar time)"という語句のみを含む。NLPデータ 例えばNL Pデータベース218は、8時(8 c l'ock)、10時(10

ベース218は以下で更に記述される。

[0045]

私の (my) " もしくは"私のもの (mine)"は話し手の氏名で置換される この工程の目的は、ユーザがNLPデータベース218内にユーザ特有の事実 (who are my children)"という文は、"誰がDeanの この場合に"Dean"は話し手の固有名称である。此処でも、この工程は簡 ブロック332にては、上記語句内の各代名詞が代名詞リプレーサ211によ を記憶してアクセスするのを許容することである。例えば"誰が私の子供ですか 子供ですか(who are Dean's children)"に変換され 単なループにより達成され得るが、該ループは、代名詞に関して上記語句を検索 し、発見された代名詞を適切な固有名称に置換する。置換が行われた語句内の箇 り適切な名称で置換される。例えば一実施例において、代名詞"私(1)"、 所を追跡する為に配列が使用され得る。

[0046]

ブロック334にては語句内の個々の単語が単語ウェイタ208により、該語 "a"、"an"、"the"および他の品詞などの単語に割当てられる。最高 えられる。例えばそれらの単語とは、NLPデータベース218内の全ての動詞 句の全体的意味に対する"重要性"もしくは"有意性"に従い加重される。例え 加重係数は、上記語句の意味に対して有意性を有すると思われる単語に対して与 NLPデータベース218に対する更 上記語句内の他の全て ば一実施例においては、割当てられる3種の加重係数が在る。最低加重係数は、 形容詞および固有名称である。中間加重係数は、 の単語に与えられる。この加重の目的は、 に強力な検索を許容することである。 、名駒、

[0047]

図4には、一実施例におけるNLPデータベース218の選択列の例が示され 12A~412Nを備えている。列400において各エントリは、上記NLPデ ・タベースに対して"既知"の語句を表している。列402においては、列40 ている。NLPデータベース218は複数の列400~410および複数の行4 0における各エントリに対して必要な単語の個数が示される。列404において

8

特数2002-524806

図4に示されたNL Pデータベース218は本発 列408および410においては、列400における各エントリに対し、一個以 は、列400における各エントリに対する関連文脈または下位文脈が示される。 明を教示すべく簡素化された例に過ぎないことを衉記されたい。他の実施例は、 **種々のエントリにより更に多くのもしくは更に少ない列を有し得る。** 上の関連動作が示される。尚、

[0048]

究所 (Cognitive Science Laboratory) により開 図3Bに戻ると、ブロック336にてNLPデータベース218は、該NLP データベース218の列400内のエントリが上記語句内の各単語(またはそれ らの同義語)のいずれかを含むか否かということと、これらの単語の相対重みと こ基づき、上記語句に対する可能的一致に関して検索される。プロック338に 語検索は当業界で公知であり、且つ、バージニア州、アーリントンのDT So ftware, Inc. による製品"dtsearch"などの市販の検索エ ンジンにより実施され得る。同様に、同義語を使用した検索は、当業界で公知で ニュージャージー、プリンストンのプリンストン大学の認識科学研 可能的一致エントリの各々に対して確信値が生成される。データベースの加重単 ては、上記語句内における各単語の発生回数と、各単語の相対重みとに基づき、 発された"WordNet"の如き公開ツールなどを使用して達成され得る。 記検索エンジンは、自然言語プロセッサ202の一体的部分とされ得る。 あり、且つ、

[0049]

判定340にて自然言語プロセッサ202は、可能的一致エントリのいずれか る。このスレッショルドTは、NLPデータベース218内の各エントリのいず れかに対して上記語句が一致したか否かに関する判定が為される為の容認可能最 コーはブロック342へと進み、選択的エラー・メッセージがユーザに対してデ **ィスプレイ104上で視覚的表示され、または、スピーカ112を介して音響的** ご再生される。一実施例においてユーザに対して表示されるエラー・メッセージ があれば、そのエラー・メッセージの形式は、最高確信NLPデータベース・エ が、一定の所定最小スレッショルドT以上である確信値を有するか否かを決定す 低確信値を表す。もしT以上の確信値を有する可能的一致エントリが無ければフ

ントリ内で何個の"適合 (hit)" (すなわち、上記語句の何個の照合対象単 語)が発見されたかに依存し得る。また、ゼロもしくは1個の適合であった場合 には、2個以上の適合だった場合とは異なる形式のエラー・メッセージが生成さ

[0000]

但しNLPデータベース218内において、当該エントリに対してT以上の確 言値が存在する少なくとも一個のエントリが在れば、フローはブロック344へ イズ"単語としては、上記語句内の他の単語と比較して該語句の全体的意味に関 してそれほど寄与しない単語が挙げられる。斯かる単語としては、冠詞、代名詞 接続詞、および同様の性質の単語が挙げられる。"非ノイズ(nonーnoi se)" 単語は、上記語句の全体的意味に関して相当に寄与する単語が挙げられ る。"非ノイズ"単語としては、動詞、名詞、形容詞、固有名称、および同様の と進み其処で上記語句からは"ノイズ (noise)" 単語が廃棄される。 性質の単語が挙げられる。

[0051]

フローは図3 Cに進みブロック346にては、最高確信一致エントリに関して NL Pデータベース218の列402から非ノイズ単語要件が読出される。例え どの映画が放映されるか (what movies are playing ば、もし最高確信照合対象語句が行412A内のエントリ(例えば、 at time)")であれば、非ノイズ単語必要個数は3である。

[0052]

NL Pデータベ ース218から読出された最高確信エントリ内に実際に存在するか否かを決定す るテストが行われる。このテストはブロック336において実施された適合形式 上記語句と容認可能に一致しなくてもエントリは最小スレッショルドTよりも大 **検索(relevance—style search)の精度の検証であり、** 判定348にては、上記語句からの非ノイズ単語必要個数が、 きな確信値を生成する可能性が在ることは理解される。

0053

判定348で実施される上記テストの性質は、ブーリアン・テスタ210によ

8

特表2002-524806

然言語プロセッサ202はアプリケーション・インタフェース220に対して列 はその同義語)内の非ノイズ単語の各々が上記最高確信エントリ内に実際に存在 る十分な個数の必要単語が在れば、フローはブロック350〜と進み、其処で自 408もしくは410からの関連動作を行うことを指示する。付加的な動作列も り実施される論理的"AND"テストである。このテストは、上記語句(もしく するか否かを決定するものである。もし上記最高確信エントリ内に実際に存在す 存在し得ることは理解されよう。

[0054]

8のブーリアン・テストが、列400内のエントリ内の語句には実際に3個の非 ノイズ単語が在ることを決定したなら、列408における関連動作(例えば、映 1")であれば関連動作は、自然言語プロセッサ202が(不図示の)テキスト /音声アプリケーションに対し、スピーカ112を介してユーザに現在時刻を音 **声発生することを指示することである。他の例として、もし最高確信エントリが** the news)")であれば、第1の関連動作はインターネット上の所定の ニュース・ウェブ・サイトにアクセスすることであり、且つ、第2の関連動作は メージを表示することを指示することであり得る。別のもしくは付加的な動作も 例えば、最高確信エントリが行412A内のエントリであり、且つ、判定34 画ウェブ・サイトにアクセスする)が行われる。上記NLPデータベースにおけ る他の各エントリは他の関連動作を有する。例えば、最高確信エントリが行41 (不図示の) イメージ表示アプリケーションに対し、上記ニュースに関連するイ 2 Eにおけるエントリ(例えば、"いま何時(what time is "ニュースを見せて下さい (show 行412Nのエントリ (例えば、 実施され得る。

[0055]

セッサ200に対し、列404の関連文脈に対して文脈固有文法212を有効化 同様に、もし最高確信エントリが判定348で決定された如き上記語句からの 非ノイズ単語必要個数を含むならば、自然言語プロセッサ202は音声認識プロ することを命令する。故に行412Aに対しては、"映画 (movies) " と いう文脈に関する文脈固有文法212が有効化される。故に、図3Aのブロック

前述の如く、適切な文脈固有 300にて音声認識プロセッサ200に対して次の発語が提供されたとき、音声 器髄プロセッサ200は一般文法214を検索する前に"映画 (movies) 文法212を有効化すると高速で好結果な音声認識の可能性が高まると共に、 話的な形式でコンピュータと対話するユーザの能力が強化される。 に対する有効化文脈固有文法212を検索する。

トリ内に上記語句からの非ノイズ単語必要個数が実際に存在しなければ、フロー **オブロック354に進み、其処でユーザはディスプレイ104もしくはスピーカ** 112を介し、最高確信エントリが表されたか否かが問われる。例えばユーザが "今日のIBM株の売りはいくらか (How much is IBM sto ck selling for today)"と発語すると、NLPデータベ **適合率は高いが、必要単語(もしくはそれらの同義語)の個数は十分でない。故** にユーザはブロック354にて、ユーザが"1998年8月28日の1BM株の 但し判定348にて、NLPデータベース218から読出された最高確信エン 1998)")を意味したか否かが問われ ース218内の最高確信エントリは行412B内のエントリである。この場合、 of IBM the price 28, 値段はいくらか (what is ck on August

[0057]

ューザは、肯定的にもしくは否定的に応答し得る。ユーザが肯定的に応答した と判定356で決定されたなら、ブロック350にては最高確信エントリに関連 する動作が行われ、且つ、ブロック352にては関連する文脈固有文法212が 有効化される。

[0058]

ログを使用して情報が問われる。例えばユーザが"XICOR株の今日の値段は 其処で、上記最高確信エントリに関してNLPデータベース218の列404か ら関連文脈が読出され、且つ、ブロック360にてユーザは文脈用対話型ダイア 但し判定356にてユーザが否定的に応答したならフローは図3Dへと進み XICOR いくらか (what is the price of

(54)

特数2002-524806

k today)"と発語すると共にNLPデータベース218からの最高確信 であったとすれば、ブロック354でユーザは、それがユーザの意図した処であ エントリは行412B (例えば"日付のIBM株の値段はいくらか (what stock of IBM price るか否かが問われる。 is the

[0059]

2 を介してユーザには株文脈に対する文脈用対話型ダイアログが呈示される。斯 かる文脈用対話型ダイアログは、XICOR株に対する名称および相場表示器記 号に関してユーザに問う段階を含み得る。ユーザは必要情報を話すことで応答し 得る。各可能的文脈に対しては、異なる文脈用対話型ダイアログが使用され得る 例えば"気象(weather)"文脈用対話型ダイアログは、気象情報が所 ニュース・サイトに対するインターネットURL、または、他の関連情報をユー ック358にて列404から読出され、ディスプレイ104およびスピーカ11 ス(n e w s)"文脈用対話型ダイアログは、記事の種類、ニュース・ソース、 もしューザが否定的に応答したなら、"株(s t o c k)"という文脈がブ 望される場所 (例えば市) の名称をユーザに問う段階を含み得る。また"ニュ ザに問う段階を含み得る。

[0900]

発見され、且つ、更なる情報をユーザに問うこと無く適切な動作が行われる。故 段文法214および文脈固有文法212はブロック362にて新情報を含むべく 上記文脈用対話型ダイアログが完了したなら、NLPデータベース218、 に本発明は、ユーザにより発語された語句を認識すべく適合的に"学習" 更新される。この様にして、ユーザは次回にてその情報を要求し、

[0061]

2、一般文法214および聴音書取文法216のひとつ以上が、各エントリに関 **連する (不図示の) タイムスタンプ値も含み得る。一致エントリが使用される毎** または、ユーザにより開始されたとき、一定の日付および時刻より以前のタイム 本発明の一実施例においては、NLPデータベース218、文脈固有文法21 に、そのエントリに関連するタイムスタンプ値は更新される。定期的間隔にて、

[0062]

ンタフェースを有するワールド・ワイド・ウェブ ("WWW" もしくは "ウェブ 本発明の一実施例において音声認識および自然言語処理は、文脈依存音声式イ ")などのネットワーク・オブジェクトと対話すべく使用され得る。

[0063]

図5は好適なダイアログ定義ファイル(DDF)500を示すが、該ファイル はウェブ・ページなどであるインターネット・オブジェクトに対して上記音声認 識および自然言語処理を関連させる為に必要な情報を表している。最も簡単な実 ーブル510から成るが、該DDFは図5に示された如く付加的な文脈固有文法 ファイル214および付加的な自然言語処理 (NLP) データベース218も含 み得る。DDF500の好適実施例は、ネットワーク・オブジェクト・テーブル 5.10と、文脈固有文法ファイル214と、自然言語処理データベース218に 500は、テキストもしくはグラフィック・ファイル、または、好適実施例にて 施例においてダイアログ定義ファイル500はネットワーク・オブジェクト・テ これらの構成要素は圧縮されると共に イル500へと結合され得る。文脈固有文法ファイル214および自然言語処理 ネットワーク・オブジェクト の動作、文法もしくはエントリに対して関連付けるメモリ・ツリー、チェーンも 、Lempe1-2iv圧縮などの当業界で公知の任意の方法によりDDFファ · テーブル510は、リソースのアドレスをNLPデータベース218内の種々 データベース218は、上述の各段落で説明した。 対するエントリを含むファイルとを含む。 しくはテーブルなどのメモリ構造である。

[0064]

ネットワーク・オブジェクト・テーブル510の好適実施例は図6に示される 図6はメモリ・テーブルを示すが、該メモリ・テーブルは次のものに対するエ ントリ列を含み得る:ネットワーク・オブジェクト520、テキスト→音声(T

(52)

特数2002-524806

スタンプ値を有する各エントリは夫々のデータベース/文法から削除される。こ これは、誤った一致の回避も促進する の様にして上記データベース/文法は古いもしくは期限経過エントリを"一掃" することで効率的サイズに保持され得る。

[0062]

ンタフェースを有するワールド・ワイド・ウェブ ("WWW" もしくは "ウェブ 本発明の一実施例において音声認識および自然言語処理は、文脈依存音声式イ ") などのネットワーク・オブジェクトと対話すべく使用され得る。

[0063]

図5は好適なダイアログ定義ファイル(DDF) 500を示すが、 該ファイル はウェブ・ページなどであるインターネット・オブジェクトに対して上記音声認 職および自然言語処理を関連させる為に必要な情報を表している。最も簡単な実 ーブル510から成るが、該DDFは図5に示された如く付加的な文脈固有文法 ファイル214および付加的な自然言語処理 (NLP) データベース218も含 4.得る。DDF500の好適実施例は、ネットワーク・オブジェクト・テーブル 5.10と、文脈固有文法ファイル214と、自然言語処理データベース218に イル500~と結合され得る。文脈固有文法ファイル214および自然言語処理 500は、テキストもしくはグラフィック・ファイル、または、好適実施例にて 施例においてダイアログ定義ファイル500はネットワーク・オブジェクト・テ 対するエントリを含むファイルとを含む。これらの構成要素は圧縮されると共に データベース218は、上述の各段落で説明した。ネットワーク・オブジェクト 、Lempe1-2iv圧縮などの当業界で公知の任意の方法によりDDFファ · テーブル510は、リソースのアドレスをNL Pデータベース218内の種々 の動作、文法もしくはエントリに対して関連付けるメモリ・ツリー、 しくはテーブルなどのメモリ構造である。

[0064]

ネットワーク・オブジェクト・テーブル510の好適実施例は図6に示される 図6はメモリ・テーブルを示すが、該メモリ・テーブルは次のものに対するエ ントリ列を含み得る:ネットワーク・オブジェクト520、テキスト→音声(T

214で置換され且つ既存ファイルは無効化されることが示される。"文法追加 フラグ528は、既存の文脈固有文法ファイルと同時に新たな文脈固有文法フ アイルが有効化されるべきことを示す

[0067]

8 内に記憶されたコマンドに対するポインタと、のいずれかとされ得る。例えば 10"であり、これは上記NLPデータベース内の第210番目のコマンドへの "いいえ"という答えならコンピュータは上記NLPデータベース し"はい/いいえ"フラグ530がマークされたなら、標準的な"はい/いいえ ンピュータに対して肯定的なコマンドを話すと、コンピュータは"はいなら行う エントリ532内に示されたコマンドを実行する。同様に、コンピュータに対 して否定的なコマンドを話すと、コンピュータは"いいえなら行う"エントリ5 3 4 内に示されたコマンドを実行する。"はいなら行う"列532および"いい 行540Bに示された如く、"はい∕いいえ"フラグがマークされる。コンピュ "はいなら行う" エン えなら行う"列534内の各エントリは、コマンドと、NLPデータベース21 "はい/いいえ"文法構造に関している。 文法が有効化される。標準的な"はい/いいえ"文法が有効化されたとき、 トリ532内の対応コマンドが実行され;この特定の場合にエントリは数字 -タに対して"はい"などの肯定的回答が与えられると、 内の第211番目のコマンドを実行する。 最後に、残りの列のエントリは、 参照である。

[0068]

のインターネットワークを行うコンピュータに対して音声および音響でコマンド 3.A乃至図3.Dの方法と組合せて使用され得る。ブロック6.0.2にて、ワールド 図7Aに行くと、ワールド・ワイド・ウェブをブラウズするコンピュータなど オブジェクト)が提供される。上記ウェブ・ブラウザはインターネット上をナビ ゲートすべく使用されるプログラムであり、当業界で公知である。ブロック60 2においてブラウザに対してURLを提供する段階は、ユーザがURLを手動タ イプ入力するか、または、選択されたウェブ・サイトURLに対する"リンク' ·ワイド·ウェブ・ブラウザに対してはウェブ・サイトURL (ネットワーク を提供する方法およびシステムが示されている。図7A乃至図7Cの方法は、

(38)

ンドの結果でもあり得る。URLが与えられたならコンピュータはブロック60 4にて、上記URL内で指定されたウェブ・サイトのインターネット・アドレス を決定し得るか否かを判定せねばならない。この決定プロセスは、当業界で公知 **鼻なければ、ブロック605にてはブラウザ画面内にエラー・メッセージが表示** アドレスが決定されるなら、コンピュータはプロック606にてウェブ・ページ この段階はまた、NLPデータ ベース218内の各エントリに関連付けられた動作に関して前述された発声コマ のプロセスである。もしコンピュータが上記インターネット・アドレスを決定し され、システムはその初期開始状態600~と戻る。もし上記インターネット・ をユーザに選択させる如く可及的に容易である。 に対する要求をウェブ・サイトへと送信する。

[6900]

は、ウェブ・ページの送信が不首尾となれば、ブロック605にてはブラウザ画 ブロック608にては、上記ウェブ・サイトがウェブ・ページを送信するか否 かに依存して判定が為される。もし上記ウェブ・サイトが応答しなければ、また 面内にエラー・メッセージが表示され、システムはその初期開始状態600へと **戻る。もし上記ウェブ・サイトがウェブ・ページを戻したなら、ブロック610** ではブラウザ画面内にそのウェブ・ページが表示される。

[0700]

判定ブロック612にてコンピュータ100は、上記ウェブ・サイトに対応す るDDFファイル500が既に該コンピュータ100内に存在するか否かを決定 する。もし上記DDFファイルが存在すれば、フローは図7Cへと進み、存在し なければフローは図7B〜と進む。

[0071]

テキスト・マークアップ言語(HTML)内にコード化されているか否かを吟味 する。(HTMLは当業界で公知であることから、該言語の詳細は本明細書中で 図7Bに行くと、もしDDFファイル500が存在しなければ上記コンピュー タは、DDFファイル500の場所がURLとしてウェブ・ページのハイパー は論じない。)HTMLコード内へのDDFファイル場所のコード化は、

DDF= "http://www.conversationa

特数2002-524806

9

lsys.com/Converselt.ddf"> などの様に初期HTMLメタ・タグ内にDDFファイル場所を列挙するか、ブラウザによりサポートされるHTMLの変種へとスクリプト・タグを直接的に書き

込むことで行われる:

<DDF= "http://www.conversationalsys
com/Converselt.ddf">

/

もしDDFファイル場所の情報がウェブ・ページ内にコード化されていれば、その場所のインターネット・アドレスはブロック616にて決定され、ブロック626にてコンピュータはDDFファイル500の転送を要求する。

[0072]

代替的に、もしDDFファイル500の場所が上記ウェブ・ページ内にコード Lされていなければ、DDFファイルの場所が記憶され得る幾つかの代替的場所 が在る。DDFファイルの場所は、ルート・ディレクトリ内の一定のファイル箇 は図1の記憶媒体108などの異なる中央箇所において、事前定義箇所内に記憶 され得る。ブロック618および620はこれらの可能性に対してテストを行う 。ブロック618は、DDFファイルが上記ウェブ・サイトに配置されているか 否かを決定する。この工程にて上記コンピュータは、DDFファイル500の存 在に関して照会を行うクエリを上記ウェブ・サイトに送信する。もし上記ウェブ ・サイトにDDFファイル500が存在すれば、上記コンピュータはブロック6 26にてDDFファイル500の転送を要求する。もしDDFファイル500が 上記ウェブ・サイトに配置されていなければ、上記コンピュータはブロック62 0 にて上記ウェブ・サイトに関するDDFファイルの存在に関して中央箇所に照 ピュータはブロック626にて上記DDFファイルの転送を要求する。もしDD オブジェクト・テーブル510、上記ウェブ・サイトに関連するNLPデータベ 所などのウェブ・サイトにおいて、または、別のインターネット・サーバもしく 会する。もし上記DDFファイルが上記ウェブ・サイトに存在すれば、上記コン Fファイル500が発見されなければ、ブロック622にては、ネットワーク

ース218、および、以前に訪れた一切のウェブ・サイトに対する文脈固有文法214、などの既存DDFファイルの既存構成要素が作動停止される。更にプロック624にて、上記ウェブ・サイトは非音声起動ウェブ・サイトとして処理されると共に、標準的文法ファイルのみが使用される。標準的文法ファイルは上記システム上に存在する文法ファイルであるが、上記ネットワーク・オブジェクトに関連する文脈固有文法ファイルと関連する一切の文法は除かれる。

[0073]

もしブロック626にてDDFファイル500が要求され乍らもその転送が不首尾であれば、ブロック622では存在する一切のDDFファイル500の一切の既存構成要素が作動停止され、且つ、ブロック624では上記ウェブ・サイトが非音声起動ウェブ・サイトとして処理され且つ標準的文法ファイルのみが使用される。

[0074]

もしブロック626にてDDFファイル500が要求されると共にその転送が好結果であれば、そのDDFファイルはブロック630にて一切の先行DDFファイルに置き換わる。ブロック632にては、ネットワーク・オブジェクト・テーブル510、文脈固有文法ファイル214およびNLPデータベース218などのDDFファイル500の一切の構成要素が抽出される。図2の機能要素を含め図3A乃至図3Dに示された方法を実現する為に必要なソフトウェアを獲得する為に、同様の技術が使用され得る。

[0075]

フローは図7 C~と移動する。ブロック634にてネットワーク・オブジェクト・テーブル510は上記コンピュータによりメモリ内に誘取られる。ブロック636により決定された如く、もし上記ウェブ・ページURLが上記サイト・ネットワーク・オブジェクト・テーブル510内に存在すれば、斯かるウェブ・ページURLは図6に示された如く上記テーブルの行540A乃至540Eにより表される。もし上記ウェブ・ページに対する非音声相互作用が存在し、処理は終了する。

00761

ブロック636で決定された如くもし上記サイト・ネットワーク・オブジェクト・テーブル510内にウェブ・ページURLが存在すれば、ブロック638にて上記コンピュータは、TTSフラグ522がマークされているか否かをチェックすることにより上記ウェブ・ページにテキスト音声524が関連付けられるか否かを決定する。もしテキスト音声524が在れば、それはブロック640にて

特級2002-524806

(35)

【図2】図1のコンピュータ・システムにおける本発明のCPU 102および記憶媒体108の拡大機能ブロック図。

【図34】コンピュータに対する対話型音声認識/自然言語処理を提供する方法を示すフローチャート。

【図3B】コンピュータに対する対話型音声認識/自然言語処理を提供する方法を示すフローチャート。

【図3C】コンピュータに対する対話型音声認識/自然言語処理を提供する方法のフローチャート。

ロック644で有効化され、次にブロック646でNLPデータベース218が有効化される。もし上記ウェブ・ページに文脈固有文法ファイルが関連付けられていなければ、ブロック646にてNLPデータベース218のみが有効化され

る。上記NLPデータベースが有効化646されたなら、上記システムは上述の

如く図3A乃至図3Cの様に挙動する。

[0077]

要約すると、本発明はコンピュータの為のネットワーク用対話型ユーザ・イン

タフェースに対する方法およびシステムを提供する。ダイアログ定義ファイルを介してインターネット・オブジェクトに連結された文脈固有文法を使用することにより、本発明は音声認識時間を短縮すると共に、ユーザが会話的な形式でウェ

種々の文法およびNLPデータベースを適合更新することにより、本発明は対

話効率を更に高める。

[0078]

ブ・ページなどのインターネット・オブジェクトと対話する能力を高める。

様に提供されたものである。当業者であればこれらの実施例の種々の改変例は容易に明らかであると共に、本明細書中で定義された包括的原理は発明能力を使用

せずとも他の実施例に適用され得る。故に本発明は本明細書中に示された各実施 例に限定されることは意図されず、本明細書中に開示された原理および新規特徴

各好適実施例に関する上記記述は、当業者が本発明を実施もしくは使用し得る

上記ウェブ・ページに関連付けられた文脈固有文法ファイルが在れば、それはブ

音声発生され、フローは継続する。もし判定プロック642にて決定された如く

【図3D】コンピュータに対する対話型音声認識/自然言語処理を提供する 方法を示すフローチャート。 【図4】本発明の好適な自然言語処理(N.L.P)データベースの選択列を示-図.

【図5】本発明に係る好適なデータベース定義ファイル (DDF) を示す図

【図6】本発明の好適なネットワーク・オブジェクト・テーブルの選択列を

【図7A】ウェブ・ページなどのネットワーク・オブジェクトに対する対話型音声認識/自然言語処理のリンクを示す、本発明の方法のフローチャート。

【図78】ウェブ・ページなどのネットワーク・オブジェクトに対する対話型音声認識/自然言語処理のリンクを示す、本発明の方法のフローチャート。

[図7C] ウェブ・ページなどのネットワーク・オブジェクトに対する対話型音声認識/自然言語処理のリンクを示す、本発明の方法のフローチャート。

【図8】インターネットを介して他のコンピュータ、記憶媒体およびウェブ・サイトに接続しているコンピュータ・システムを示す図。

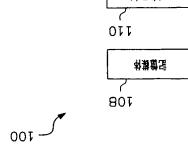
に一致する最大範囲に従うものとする。

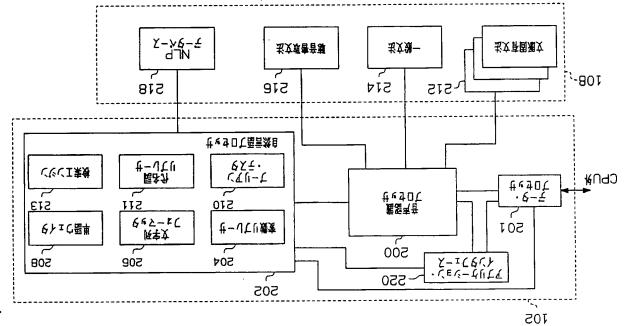
[図面の簡単な説明]

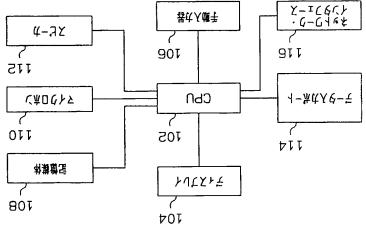
[図1] 本発明と共に使用される好適なコンピュータ・システムの機能ブロ

-32-

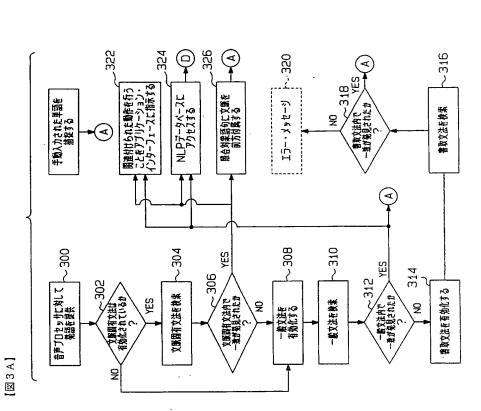
[図]







 \bigcirc



 \sim 332

代名詞置換

関連付けられた ワイルドカード機能により 単語変数を置換する

NLP分析の為に 語句をフォーマットする \sim 334

個々の単語を加重する

-36-

-35-

エラー・メッセージ

旲

いずれがの一致が 疑信値2Tを 有するか?

~340

√_{YES} 344

語句から。ノイズ。 単語を廃棄する

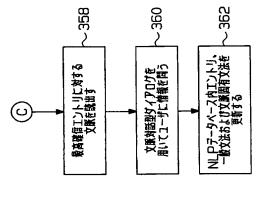
(B)

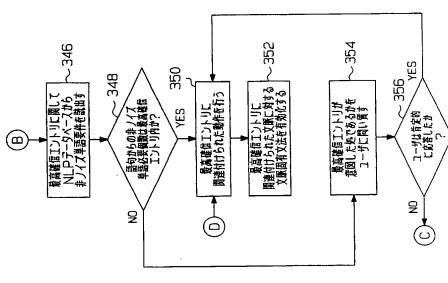
 \sim 338

可能的一致に対して 確信値を生成する

NLPデータベース を検索する [🖾 3 D]







_1
(
200

X->6-₹dJN

518

[図2]

9
0
œ
4
2
S
1
2
0
0
7
特我

1	•	
ı		
:	١	١

[図4]

]]
ピーメ <i>ト</i> 示表き	617367347# 617367347#	(NEMS)	2	いち不丁も見きたしここ (SHOW AHT 3M WOH2)	~ N217
イスキモ き 候却 るをコギ音←	16 生生差	時 (ATME)	5	報所まり ・ (TI SI ∃MIT TAHW)	41SE ~
\$3 <u></u> 5%	イト世・てェウ集成 るを入す <i>で</i> でコ	緊 (MEATHER)	ε	本でうりにおける気はなります。 WHAT IS THE WEATHER IN \$LOCATION)	4120
1944	インせ・てュウ[F産業 るを入すででコ	(STOCKS)	Þ	(SELL IBM STOCK AT \$DOLLARS)	415C ~
19世星	イト世・てェ ご語券 るを入生 <i>で</i> でコ	(STOCKS)	₽	日付(今)のIBM等の権限はいくらか (WHAT 15 THE PRICE OF IBM STOCK ON \$DATA)	4128
1942	イトせ・てェウ画 点 るを入すぐでコ	新国 (WOVIES)	3	なるがち突然ない変化のとは)(ま)間も BAIYAJU BAA SELV MOTES TA (1)MIT TA	~ ASIA
S計値	1. 計0種	潮文並不入潮文	器車菱沙	9蜀	
۲	201		705	205	₹
014	804	404	405	004	218

514

-40-

ر 10

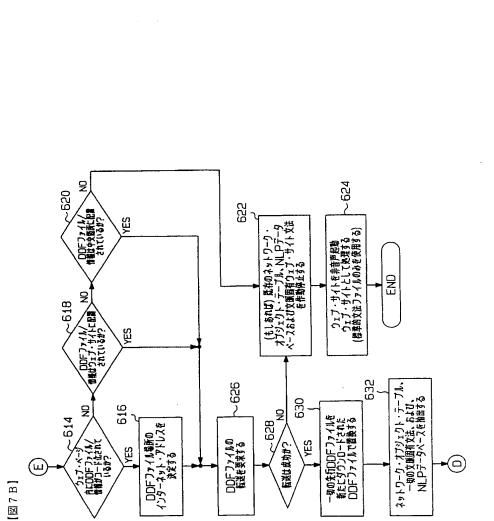
-39-

		ラウザ画面に ・・メッセージが ~60E 表示される					
009	### 7ラウザに対して ウェブ・サイトURL を提供する	604 プラウ・インターネット・アドレス NO Iラー・メ 表表に得るか? ves	ウェブ・サイトにウェブ・ ページ要求を送信 ページ要求を送信	ウェブ・サイトは つェブ・ページを戻したか ア	▼ 7ラウザ南面にウェブ・ ページを表示する	ANDERSON FOR THE NO E ARE SOR	AES VES

		_					• •	
513	S24	×					HTTP: \\ WWW.CONVERSATIONALSYS.COM\ SEARCH	240E
					(AIV AOICE I2).	×	HTTP:\\ WWW.CONVERSATIONALSYS.COM\ PRDDUCTS/VIAVOICE	~0079
				×	・・・ TAM語音数自の方き、 (ALL NATUTAL LANGUAGE (・・・ BAA STOUGOR9	×	HTTP:\\ WWW.CONVERSATIONALSYS.COM\ PRODUCTS	240C
511	510	×			、 なまないないないになって、 (MOULD YOU LIKE TO LEARN ABOUT NEW PRODUCIS?) ABOUT NEW PRODUCIS?)	×	HITP:\\ WWW.CONVERSATIONALSYS.COM\ NEWS	~8079
			×		CONVERSATIONAL SYSTEMS.). \$354 (HELLO, WELCOME TO CONVERSATIONAL SYSTEMS^	×	HTTP:// WWW.CONVERSATIONALSYS.COM	~ 4042
रेक्स्याग र्हे	रुक्षण्य रुति	2¥1111 /11#	松	轍	有音イス‡モ	גבר 11	1018CK·0-0164]
₽ £9	283	530	9Z9 - C	929 25e	P25	2SS 2SS	250	7
V C 9	223	089	963	303	, 03			Òti

[図7C]





운

上記ウェブ・ページ に関連付けられた音声文言 はあるか?

 \sim 638

YES

640

YES

音声文言を再生する

644

YES

文脈固有文法を有効化する

NPL データベースを 有効化する

646 →

~ 642

上記ウェブ・ページ に関連付けられた文脈固有 文法はあるか?

2

サイト・ネットワーク・ オブジェクト・テーブル内に ウェブ・ページURLは 存在するか?

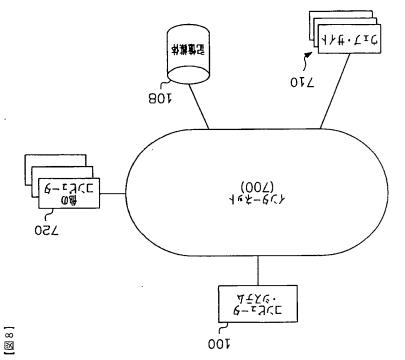
 ~ 636

ネットワーク・オブジェクト ・テーブルを競取る

-44-

-43-

. (45)



[国際調査報告]

INTERNATIONAL SEARCH REPORT

		PCT/US 99/20447	4
15g	1 pc 7 d10115/26 404129/06 806F17/27 G06F3/16	/16	
According	coording to Intermetend Peters Charaltection (IPQ or to both protocol classification and IPQ		
	A riskow szakratego Martin den mertán mesébb (shembodon system telemed by cheeftesfon symbols) 1907 7 2:11 10041 AAST		
Documents	and of a control of the fiber of the fiber of the control of the control of the control of the fibrical of the folder sections.	nchated in the faiths assected	
Backroki	Ekstrack chis have corestind dafry the besending eventh (mans of shink have met, where procless, weardt harm uned	lost, search terms used)	
o. Bocane	о оослагить сомпоелез то ше пед вумит		
Cethopony *	Cladics of decement, with technism, where appropriate, of the relevant passenges	Referent to	ert to dain like
*	EP 0 854 416 A (TEXAS INSTRUMENTS INC) 22 July 1998 (1999-07-22) page 2, 11ns 7 - 11ns 23 page 2, 11ns 36 - 11ns 42 page 5, 11ns 36 - 11ns 46 page 6, 11ns 20 - 11ns 45 page 6, 11ns 20 - 11ns 45 page 6, 11ns 20 - 11ns 45	1-3,7-11	
> -		12-14	
×	EP 0 747 881 A (AT & T CORP) 11 December 1996 (1996-12-11) column 4, Ifna 12 - Ifna 27 column 5, Ifna 12 - Ifna 24 column 6, Ifna 18 - Ifna 28	1,7,8	
>-	 	4-6, 12-14	
	/-		
<u>}</u>	e constituedan of bas C.	Point forth members are lefted in errex.	
"A" document		The terror control published where the returnational time data or plointy date and not be confied with the supplication but other the understand the published or heavy underlying the	
E' eullard Alling da 1' decume	*	injuries; document of particular reference; the distinct hiverflor document for correlative forms or carerial be consistent to havine an investigate darp when the draument is these alone havine.	
o de	with Budden by this dailed in the delay of many of grantees of gra	decurrent of pushtader reforement the delatived interestion countries for scattering the factories of the state of scattering the scatteri	
Date of fre	of the extent completion of the international nearth	Date of mailing of the international search report	
=	15 December 1999 11/01/2000	,2000	
Nome and m	Anthogon of the ISA N Zoon John Clan, P. S. Sol of Fluoritann ? N Zoon John Chault N Zoon John Chault N Soon John Chault N Soon John Chault Sec (sel10) 940-9016	nteu ofter Ranos Sánchez, U	

--46--

page 1 of 2

特級2002-524806

(41)

-48-

A E D 654 417 A (TEXA STRUMENTS INC) A E D 654 417 A (TEXA STRUMENTS INC) 2 2 daly 1998. (1996-07-22) column 5, 11ne 5 - column 4, 11ne 2 column 5, 11ne 1 - 11ne 14 column 5, 11ne 2 - 11ne 37 A DE 44 40 598 C (51996-07-23) A DE 44 40 598 C (51996-07-23) A DE 44 40 598 C (51996-07-23) A W 96 35491 A (2018870N ROBERT DENIS 1,2,7-9 1,8,11154 1598 (1996-08-13) column 4, 11ne 26 - column 5, 11ne 8 column 4, 11ne 26 - column 5, 11ne 8			Int Servet Application No PCT/US 99/20447
EXAS INSTRUMENTS INC) 8-07-22) 8-07-22) 1-10-14 - 110-37 - 110-37 - 10-23) - 10-23 - 10-23 - 10-23 - 10-23 - 10-23 - 10-23 - 10-23 - 10-23 - 10-23 - 10-23 - 20-23 -	4	C.(Continuation) DOCUMENTS COMMITTED TO BE ARE EVANT	
EXAS INSTRUMENTS INC) 8-07-22) -6-07-22) -11ne 14 -11ne 37 -11ne 37 -11ne 37 -11ne 37 -12ne 38 -12ne 38 -12ne 38 -12ne 39 -02-23 -02-23 -02-23 -03-23	3	den of decembers, with indexelventations appropriate, of the missian perseages	Referent to clabs No.
TEPENS A6) -05-23) -colum 3, 11ne 17 -colum 5, 11ne 8 -colum 5, 11ne 8		4 417 A (TEXAS INSTRUMENTS 1998 (1998-07-22) 3, line 36 -column 4, line 5, line 1 - line 14 5, line 29 - line 37	1-14
-column 3, 11ne 17 -column 5, 11ne 8 -column 5, 11ne 8		DE 44 40 598 C (SIENENS A6) 23 Nay 1996 (1996-05-23)	1,2,7-9
49870W ROBERT DEWIS 988-08-13) -column 5, 11ne 8		abstract column 2, line 46 ∼column 3, line 17	
		WO 98 35491 A (JUHNSTON ROBERT DENIS BRITISH TELECORM (BB) 13 August 1998 (1998-08-13) column 4, line 28 -column 5, line 8	1,2,7-9
			~~
		,	

INTERNATIONAL SRARCH REPORT

page 2 of 2

フロントページの続き

(8) 指定国 EP(AT, BE, CH, CY, DE, DK, ES, F1, FR, GB, GR, 1E, 1 T, LU, MC, NL, PT, SE), OA(BF, B J, CF, CG, C1, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG), AP(GH, GM, K E, LS, MW, SD, SL, SZ, UG, ZW), E A(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, F1, G B, GD, GE, GH, GM, HR, HU, 1D, 1L, 1N, 1S, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MN, NO, NZ, PL, PT, RD, RD, SE, SG, S1, SK, SL, TJ, TT, UA, UG, US, UZ, V

N. YU, ZA, ZW 【要約の機き】

43.